ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

Proceedings of the
13th European Conference on Cyber Warfare and Security
The University of Piraeus
Greece
3-4 July 2014

Edited by
Andrew Liaropoulos and George Tsihrintzis

acpi

A conference managed by ACPI, UK

Proceedings of the 13th European Conference on Cyber Warfare and Security ECCWS-2014

The University of Piraeus
Piraeus, Greece
3-4 July 2014

Edited by
Andrew Liaropoulos
and
George Tsihrintzis

Copyright The Authors, 2014. All Rights Reserved.

No reproduction, copy or transmission may be made without written permission from the individual authors.

Papers have been double-blind peer reviewed before final submission to the conference. Initially, paper abstracts were read and selected by the conference panel for submission as possible papers for the conference.

Many thanks to the reviewers who helped ensure the quality of the full papers.

These Conference Proceedings have been submitted to Thomson ISI for indexing.

Further copies of this book and previous year's proceedings can be purchased from http://academic-bookshop.com

E-Book ISBN: 978-1-910309-25-4
E-Book ISSN: 2048-8610
Book version ISBN: 978-1-910309-24-7
Book Version ISSN: 2048-8602
CD Version ISBN: 978-1-910309-26-1
CD Version ISSN: 2048-8629

Published by Academic Conferences and Publishing International Limited
Reading
UK
44-118-972-4148
www.academic-publishing.org

Contents

Paper Title	Author(s)	Page No.
Preface		iii
Committee		iv
Biographies		vi
Possible Scenarios and Maneuvers for Cyber Operational Area	Ugur Akyazi	1
Digital Forensics as a Science in Higher Education	Olga Angelopoulou and Stilianos Vidalis	8
Knowledge Accessibility and Cyber Macht	Leigh Armistead and Scott Starsman	17
A Strategic Approach to Managing Security in SCADA Systems	Mehdi Asgarkhani and Elena Sitnikova	23
Fuzzy Application With Expert System for Conducting Information Security Risk Analysis	Jiří Bartoš, Bogdan Walek, Cyril Klimeš and Radim Farana	33
Identity Multipliers and the Mistaken Twittering of 'Birds of a Feather'	David Cook	42
Secret Sharing Framework Based on Digital Certificates	Paul Crocker and Adolfo Peixinho	49
Improving Cyber-Security Awareness on Industrial Control Systems: The CockpitCI Approach	Tiago Cruz, Jorge Proença, Paulo Simões, Matthieu Aubigny, Moussa Ouedraogo, Antonio Graziano and Lasith Yasakhetu	59
Putting Counterintelligence in Cyber Counterintelligence: Back to the Future	Petrus Duvenage and Sebastian von Solms	70
Information Security Economics: Induced Risks and Latent Costs	Evangelos Frangopoulos, Mariki Eloff and Lucas Venter	80
The Opportunities of National Cyber Strategy and Social Media	Arto Hirvelä, Aki-Mauri Huhtinen and Tommi Kangasmaa	88
Online Social Networks: A Vehicle for Malware Propagation	Ehinome Ikhalia and Johnnes Arreymbi	95
Countering Threats - a Comprehensive Model for Utilization of Social Media for Security and law Enforcement Authorities	Margarita Jaitner and Harry Kantola	102
Potential Cyber Warfare Capabilities of Major Technology Vendors	Audun Jøsang	110
Manpower Planning and Management in Cyber Defense	Ilker Kilaz, Akif Onder and Murat Yanik	116
The Effectiveness of Online Gaming as Part of a Security Awareness Program	William Aubrey Labuschagne and Mariki Eloff	125
Cyberconflict and Theoretical Paradigms: Current Trends and Future Challenges in the Literature	Andrew Liaropoulos	133
Planning Method of Information Security for Military Organizations	José Martins, Henrique dos Santos, Mendes Dias and José Borges	140
Comparison of two Specifications to Fulfill Security Control Objectives	Riku Nykänen andTommi Kärkkäinen	150
Challenges in Information Security Protection	Teresa Pereira and Henrique Santos	160
Enrolment Time as a Requirement for Face Recognition Biometric Systems	Vítor Sá, Sérgio Magalhães and Henrique Santos	167

Paper Title	Author(s)	Page No.
Retaining Control Over Private Virtual Machines Hosted by a Cloud Provider Using Mandatory Access Control, Trusted Boot and Attestation	Armin Simma and Philipp Rusch	172
Cyber Security and Civil Engagement: Case of Lithuanian Virtual Community Projects	Aelita Skaržauskienė, Agnė Tvaronavičienė and Gintarė Paražinskaitė	181
Determination of Meme Proliferation Factors	Namosha Veerasamy and William Aubrey Labuschagne	188
Integration of a Network Aware Traffic Generation Device Into a Computer Network Emulation Platform	Suné von Solms and Schalk Peach	198
Legal Solutions to State-Level Cyber Intrusion Under International law: A Maze of Legal Uncertainty or not?	Murdoch Watney	206
An Annotated Bibliographical Survey on Cyber Intelligence for Cyber Intelligence Officers	Cagatay Yucel and Ahmet Koltuksuz	213
PHD Research papers		221
Intrusion Detection System Using Bayesian Network Modeling	Chaminda Alocious, Nasser Abouzakhar, Hannan Xiao and Bruce Christianson	223
A Near-Miss Management System to Facilitate Forensic Investigation of Software Failures	Madeleine Bihina Bella, Jan Eloff and Martin Olivier	233
Requirements for Preparing the Cloud to Become Ready for Digital Forensic Investigation	Moses Dlamini, Hein Venter, Jan Eloff and Mariki Eloff	242
A Generic Framework for Enhancing the Quality Digital Evidence Reports	Nickson Karie and Hein Venter	251
A Framework to Address Challenges Encountered When Designing a Cyber-Range	Brendan Lawless, Jason Flood and Anthony Keane	258
A Cyber Attack Evaluation Methodology	Kosmas Pipyros, Lilian Mitrou, Dimitris Gritzalis and Theodore Apostolopoulos	264
The Changing Character of war in the Global Information Age	Anthimos Alexandros Tsirigotis	271
Masters Research papers		279
Spymasters Tools: A Comparative Approach to Side Channel Attacks	Tuğçe Kalkavan	281
Models for the Forensic Monitoring of Cloud Virtual Machines	Dirk Ras and Hein Venter	290
Non Academic paper		301
Distinguishing Cyber Espionage Activity to Prioritize Threats	John Hultquist	303
Work In Progress paper		309
NATO Article Statue 5 in Terms of a Cyber-War	Selcuk Dal and Kadir Ozyurt	311

Preface

This year sees the 13th European Conference on Cyber Warfare and Security (ECCWS 2014), which is hosted by The University of Piraeus, Greece.

The Conference Chair is Andrew Liaropoulos, from the University of Piraeus, Piraeus, Greece. The Programme Chair is George Tsihrintzis, University of Piraeus, Piraeus, Greece.

The Conference continues to bring together individuals working in the area of cyberwar and cyber security in order to share knowledge and develop new ideas with their peers. The range of papers presented at the Conference will ensure two days of interesting discussions. The topics covered this year illustrate the depth of the information operations' research area, with the subject matter ranging from the highly technical to the more strategic visions of the use and influence of information.

The opening keynote is given by Prof. Dimitris Gritzalis, Director of the Information Security & Critical Infrastructure Protection (InfoSec) Laboratory, Athens University of Economics and Business, Greece on the topic of "Open-Source Intelligence produced from Social Media: A proactive Cyber Defense tool" and the second day keynote will be presented by Prof. Nikolaos Bourbakis, IEEE Fellow, Director of the Assistive Technologies Research Center (ATRC) Wright State University, USA, on the topic of "Cyber-Security Challenges in the Cyber-Space".

With an initial submission of 71 abstracts, after the double blind, peer review process there are 27 Research papers, 7 PHD Research papers, 2 Masters Research papers, 1 Non Academic paper and 1 Work in Progress Paper published in these Conference Proceedings. These papers come from all parts of the globe including Australia, Austria, Czech Republic, Finland, Greece, Lithuania, New Zealand, Nigeria, Norway, Portugal, Republic of South Africa, Turkey, UK and USA.

I wish you a most interesting conference and an enjoyable stay in Greece.

Andrew Liaropoulos and George Tsihrintzis,
University of Piraeus, Greece

June 2014

Conference Committee

Conference Executive
Dr Andrew N. Liaropoulos Department of International and European Studies, University of Piraeus, Greece
Dr George A. Tsihrintzis University of Piraeus, Greece

Mini track chairs
Dr. Nasser Abouzakhar, University of Hertfordshire, UK
Dr Ahmet Koltuksuz, Yasar University , Turkey
Dr. Rupali Jeswal, Xiphos-ISS (Intelligence & Security Solutions)
Asst Prof Dr Selma Tekir, Izmir Institute of Technology, Turkey
Asst Prof Dr Serap Sahin, Izmir Institute of Technology, Turkey
Dr Martti Lehto, University of Jyväskylä, Finland

Committee Members
The conference programme committee consists of key individuals from countries around the world working and researching in the Information Warfare and Security community. The following have confirmed their participation:

Dr. Mohd Faizal Abdollah (University Technical Malaysia Melaka, Melaka); Dr. Nasser Abouzakhar (University of Hertfordshire, UK); Dr. Kari Alenius (University of Oulu, Finland); Prof. Antonios Andreatos (Hellenic Air Force Academy, Greece); Dr. Olga Angelopoulou (University of Derby, UK); Dr. Leigh Armistead (Edith Cowan University, Australia); Colin Armstrong (Curtin University, Australia, Australia); Johnnes Arreymbi (University of East London, UK); Debi Ashenden (Cranfield University, Shrivenham, UK); Dr. Darya Bazarkina (Sholokhov Moscow State Humanitarian University, Russian Federation); Laurent Beaudoin (ESIEA, Laval, France); Ass Prof. Maumita Bhattacharya (Charles Sturt University, Australia); Prof. Matt Bishop (University of California at Davis, USA); Andrew Blyth (University of Glamorgan, UK); Colonel (ret) Colin Brand (Graduate School of Business Leadership, South Africa); Dr. Svet Braynov (University of Illinois at Springfield, USA); Prof. Larisa Breton (University of the District of Columbia, USA); Bill Buchanen (Napier University, UK); Dr. Joobin Choobineh (Texas A&M University, USA); Dr. Maura Conway (Dublin City University, Ireland); Dr. Paul Crocker (Universidade de Beira Interior, Portugal); Dr. Christian Czosseck (CERT Bundeswehr (German Armed Forces CERT), Germany); Geoffrey Darnton (Bournemouth University, UK); Josef Demergis (University of Macedonia, Greece); Moses Dlamini (SAP Research Pretoria, South Africa); Paul Dowland (University of Plymouth, UK); Marios Efthymiopoulos (Political Science Department University of Cyprus, Cyprus); Dr. Ramzi El-Haddadeh (Brunel University, UK); Daniel Eng (C-PISA/HTCIA, China); Prof. Dr. Alptekin Erkollar (ETCOP, Austria); Prof. Robert Erra (ESIEA PARIS, France); John Fawcett (University of Cambridge, UK); Prof. Eric Filiol (Ecole Supérieure en Informatique, Electronique et Automatique, France); Dr. Chris Flaherty (University of New South Wales, Australia); Prof. Steve Furnell (University of Plymouth, UK); Assoc. Professor Javier Garci'a Villalba (Universidad Complutense de Madrid, Spain); Kevin Gleason (KMG Consulting, MA, USA); Dr. Michael Grimaila (Air Force Institute of Technology, USA); Prof. Stefanos Gritzalis (University of the Aegean, Greece); Dr. Marja Harmanmaa (University of Helsinki, Finland); Ulrike Hugl (University of Innsbruck, Austria); Aki Huhtinen (National Defence College, Finland); Bill Hutchinson (Edith Cowan University, Australia); Dr. Berg Hyacinthe (State University of Haiti, Haiti); Dr. Abhaya Induruwa (Canterbury Christ Church University, UK); Hamid Jahankhani (University of East London, UK); Dr. Amit Jain (BenefitFocus Inc, USA); Dr. Helge Janicke (De Montfort University, UK); Joey Jansen van Vuuren (CSIR, South Africa); Saara Jantunen (University of Helsinki, Finland); Andy Jones (BT, UK); James Joshi (University of Pittsburgh, USA); Nor Badrul Anuar Jumaat (University of Malaya, Malaysia); Maria Karyda (University of the Aegean, Greece); Ass Prof. Vasilis Katos (Democritus University of Thrace, Greece); Dr. Anthony Keane (Institute of Technology Blanchardstown, Dublin, Ireland); Jyri Kivimaa (Cooperative Cyber Defence and Centre of Excellence, Tallinn, Estonia); Dr. Spyros Kokolakis (University of the Aegean, Greece); Prof. Ahmet Koltuksuz (Yasar University, Dept. of Comp. Eng., Turkey); Theodoros Kostis (Hellenic Army Academy, Greece); Prashant Krishnamurthy (University of Pittsburgh, USA); Dan Kuehl (National Defense University, Washington DC, USA); Peter Kunz (DiamlerChysler, Germany); Pertti Kuokkanen (Finnish Defence Forces, Finland); Dr. Erikk Kurkinen (University of Jyväskylä, Finland); Takakazu Kurokawa (National Defence Acadamy, Japan); Rauno Kuusisto (Finnish Defence Force, Finland); Tuija Kuusisto (National Defence University, Finland); Dr. Laouamer Lamri (Al Qassim University and European University of Brittany, Saudi Arabia); Michael Lavine (John Hopkins University's Information Security Institute, USA); Martti Lehto (National Defence University, Finland); Tara Leweling (Naval Postgraduate School, Pacific Grove, USA); Paul Lewis (technology strategy board, UK); Peeter Lorents (CCD COE, Tallinn, Estonia); Hossein Malekinezhad (Islamic Azad University,Naragh Branch, Iran); Mario Marques Freire (University of Beira Interior, Covilhã, Portugal); Ioannis Mavridis (University of Macedonia, Greece); Rob McCusker (Teeside University, Middlesborough, UK); Jean-Pierre Molton Michel (Ministry of Agriculture, Haiti); Durgesh Mishra (Acropolis Institute of Technology and Research, India); Dr. Yonathan Mizrachi (University of Haifa, Israel, Israel); Edmundo Monteiro (University of Coimbra, Portugal); Evangelos Moustakas (Middlesex University, London, UK); Dr. Kara Nance (University of Alaska Fairbanks, USA); Muhammad Naveed (IQRA University Peshawar, Pakistan, Pakistan); Mzukisi Njotini (University of South Africa, South Africa); Rain Ottis (Cooperative Cyber Defence Centre of Excellence, Estonia); Tim Parsons (Selex Communications, UK); Michael Pilgermann (University of Glamorgan, UK); Engur Pisirici (govermental - independent, Turkey); Dr. Muttukrishnan Rajarajan (City University London, UK); Andrea Rigoni (Booz & Company,, USA); Dr. Neil Rowe (US Naval

Postgraduate School, Monterey, USA); Raphael Rues (DigiComp Academy, Switzerland); Filipe Sa Soares (University of Minho, Portugal); Prof. Henrique Santos (University of Minho, Portugal); Prof. Chaudhary Imran Sarwar (Mixed Reality University, Pakistan); Dr. Damien Sauveron (Mathematics and Computer Sciences, University of Limoges, France); Sameer Saxena (IAHS Academy, Mahindra Special Services Group , India); Prof. Dr. Richard Sethmann (University of Applied Sciences Bremen, Germany); Dr. Yilun Shang (Singapore University of Technology and Design, Singapore); Prof. Paulo Simoes (University of Coimbra, Portugal); Prof. Jill Slay (University of South Australia, Australia); Anna Squicciarini (University of Milano, Italy); Iain Sutherland (University of Glamorgan, Wales, UK); Jonas Svava Iversen (Danish Broadcast Corporation, Denmark); Anna-Maria Talihärm (Tartu University, Estonia); Dr. Selma Tekir (Izmir Institute of Technology, Turkey); Prof. Sérgio Tenreiro de Magalhães (Universidade Católica Portuguesa, Portugal); Prof. Dr. Peter Trommler (Georg Simon Ohm University Nuremberg, Germany); Craig Valli (Edith Cowan UniversitY, Australia); Rudi Vansnick (Internet Society, Belgium); Richard Vaughan (General Dynamics UK Ltd, UK); Stilianos Vidalis (Newport Business School, Newport, UK); Dr. Natarajan Vijayarangan (Tata Consultancy Services Ltd, India); Marja Vuorinen (University of Helsinki, Finland); Michael Walker (Vodafone, UK); Prof Mat Warren (Deakin University, Australia, Australia); Dr. Kenneth Webb (Edith Cowan University , Australia); Dr. Stuart Weinstein (University of Hertfordshire School of Law, UK); Dr. Santoso Wibowo (Central Queensland University, Australia); Dr. Trish Williams (Edith Cowan University, Australia); Simos Xenitellis (Royal Holloway University, London, UK); Dr. Omar Zakaria (National Defence University of Malaysia, Malaysia);

Biographies

Conference Chair

Dr Andrew N. Liaropoulos is a Lecturer in University of Piraeus, Department of International and European Studies, Greece. He also teaches in the Joint Staff War College, the Joint Military Intelligence College, the National Security College, the Air War College and the Naval Staff Command College. He earned his Master's Degree in Intelligence and Strategic Studies at Aberystwyth University and his Doctorate Diploma at Swansea University. His research interests include international security, intelligence reform, strategy, military transformation, cyber security and Greek security policy. Dr. Liaropoulos is also a senior analyst in the Research Institute for European and American Studies (RIEAS), the assistant editor of the Journal of Mediterranean and Balkan Intelligence (JMBI) and the convenor of the Greek Politics Specialist Group (GPSG)

Programme Chair

Dr George A. Tsihrintzis received the Diploma of Electrical Engineer from the National Technical University of Athens, Greece (with honors) and the M.Sc. and Ph.D. degrees in Electrical Engineering from Northeastern University, USA. He is currently a Member of the Council, Professor and Director of Graduate Studies in "Advanced Computing and Informatics Systems" in the University of Piraeus, Greece. His current research interests include Pattern Recognition, Decision Theory, and Statistical Signal Processing and their applications in Multimedia Interactive Services, Security, User Modeling, Knowledge-based Software Systems, Human-Computer Interaction and Information Retrieval. He has (co)-authored 5 monographs, 11 edited volumes and over 250 research publications in international journals, book chapters, and conference proceedings in his areas of research interest.

Keynote Speakers

Dr. Dimitris Gritzalis is a Professor of ICT Security and the Director of the Information Security and Critical Infrastructure Protection Research Group (www.cis.aueb.gr) with the Dept. of Informatics of the Athens University of Economics and Business, Greece. He holds a B.Sc. (Mathematics, Univ. of Patras), a M.Sc. (Computer Science, City University of New York), and a Ph.D. (Critical Information Systems Security, Univ. of the Aegean). He has served as Associate Commissioner of the Greek Data Protection Commission and as the President of the Greek Computer Society. He is the representative of Greece to IFIP TC-11. For more than twenty five years he has been participating in numerous R&D and consulting projects funded by the European Commission, international organizations (EUROPOL, CEN), public administration (Ministries of the State, Government, Economy, Healthcare, Interior, Education, Social Security, etc.), private institutions (Banking, Transportation, Telecommunication, Health, Energy, etc.). His technical publications include 9 books and more than 150 papers. His current research interests focus on security in ambient intelligence, new security paradigms, critical infrastructure protection, and strategies for security-critical infrastructures.

Dr.Nikolaos Bourbakis (IEEE Fellow) received his BS degree in mathematics from the National University of Athens, Greece and his PhD in computer science and computer engineering from the University of Patras, Greece in 1963. He is currently a Distinguished Professor in Informatics & Technology, J.A Professor of Geriatrics and the Director of Assistive Technology Research Center (ATRC) at Wright State University, Ohio. He has published more than 380 articles in refereed International Journals, book chapters and Conference Proceedings and 10 books as an author or co-author editor

Mini Track Chair

Dr Ahmet Koltuksuz received his Ph.D. in the field of cryptanalysis of Symmetrical Ciphers. He is currently affiliated with Yasar University, School of Engineering Department of Computer Engineering in Turkey. His research interests are Cryptology, Theory of Numbers, Information Theory, Theory of Computation, Operating Systems, Multicore Architectures, Cyberspace Defense & Security, Open Sources Intelligence Analysis and of Computer Forensics. He is a frequent speaker on matters concerning Cyberspace Security and over the years he has been an advisor to governmental and/or nongovernmental institutions.

Biographies of Presenting Authors

Ugur Akyazi received MSc of Systems Engineering from U.S. Air Force Institute of Technology and PhD of Computer Engineering from Istanbul Technical University. After working as an instructor in Turkish Air Force Academy for nine years, he graduated from Turkish Air War College in 2013. He still works in the Network Administration Department of Headquarters.

Chaminda Alocious, is currently living and studying in England. She completed her Msc in Secure Computer Systems from University of Hertfordshire. She is a PhD research student at the University of Hertfordshire. Her primary research areas are computer network security, specially wireless MAC layer security and machines learning techniques for network intrusion detection.

Olga Angelopoulou, BSc, MSc, PhD is a lecturer and the programme leader for the MSc Computer Forensic Investigation at the University of Derby. She obtained a doctorate in Computing with the title: 'Analysis of Digital Evidence in Identity Theft Investigations' from the University of Glamorgan. Her research interests include Digital Forensics, Identity Theft, Online Fraud, Digital Investigation Methodologies and Online Social Networking.

Leigh Armistead is the President of Peregrine Technical Solutions, a cyber security company. His PhD (Edith Cowan University) focused on IW, he is Chief Editor for the *Journal of International Warfare*, the Vice-chair WG 9.10 - ICT Uses in Peace and War, on the ECCWS Editorial Review Board and is the ICCWS Programme Director.

Madeleine Bihina Bella is doing her PhD in digital forensics at the University of Pretoria in South Africa. She has received several local and international awards for her doctorate research. She is currently appointed as a Research Associate at SAP Research in Pretoria and worked previously as an IT auditor and a business analyst.

David Cook is a lecturer in the School of Computer and Security Science at Edith Cowan University in Australia. He is a researcher in the Security Research Institute ECUSRI, and specialises in e-Governance, New Media, and Information Warfare. He is the Chair of the Australian Computer Society in Western Australia.

Paul Crocker has a PhD in Mathematics from the University of Leeds, UK. He is currently at the Reliable and Secure Computation Group of the Computer Science Department at the University of Beira Interior, Portugal and member of the Portuguese Institute of Telecommunications. His research and teaching interest are Security, Parallel Computing, and Operating Systems.

Tiago Cruz is Invited Assistant Professor at the University of Coimbra (Portugal), from where he got his PhD in Informatics Engineering, in 2012. He is also a researcher at the Centre for Informatics and Systems of the University of Coimbra. His research interests include (but are not restricted to) Broadband Network Architectures, Systems, Network Management, Security and Embedded Systems Design and Critical Infrastructure Security.

Selçuk Dal received the B.S.degree in System Engineering from Turkish Military Academy in 2001. He was accepted in Naval Postgraduate School Monterey/California and received M.S. degree in Manpower Systems Analysis in 2006. He is currently studying at Turkish Army War College. He is interested in cyber warfare, cyber security, battlefield digitization and active cyber defense.

Moses Dlamini received his BSc Computer Science and Mathematics from the University of Swaziland. He received his BSc Honours and MSc Computer Science from the University of Pretoria. He is enrolled for PhD at the University of Pretoria. He has presented research papers at national and International conferences.

Petrus Duvenage ('Beer') is a counterintelligence specialist with more than two decades experience in various aspects of this field. In the course of his career, he served as an officer in the South African Defense Force, the National Intelligence Service and the State Security Agency. He has published several articles and contributed to books on counterintelligence.

Jan Eloff is the Research Director of SAP Research Pretoria where he conducts research on mobile solutions for emerging economies. He holds a PhD in Computer Science and is appointed as an Extraordinary Professor in Computer Science at the University of Pretoria. He is an expert advisor on information security to both academia and industry.

Jason flood, MSC, currently works as a security architect at IBM. He is the founder of the Irish Honeynet Project and is a board member of Irish Reporting and Information Security Service and the Irish Cloud Security Alliance. He is currently completing a PHD at the Institute of Technology Blanchardstown.

Evamgelos D. Frangopoulos holds M.Sc. Degrees in Electrical Engineering and Information Systems. He is an Information Security Professional and is also pursuing a Ph.D. at UNISA. Professor M.M. Eloff is currently with the Institute for Corporate

Citizenship, College of Economic and Management Sciences, UNISA. Professor L.M. Venter is Institutional Director of Research Support and Extraordinary Professor of Computer Science and Information Systems at North-West University.

Danlami Gabi, obtained his BSc in computer science, in 2006 from UDUSok, Nigeria. He later went to University of East London, UK, in 2011 where he obtained MSc in Information Security and Computer forensics. He is currently a lecturer at Department of Computer Science, Kebbi state University, Nigeria.

Arto Hirvelä is an Head of Research Group at the Finnish National Defence University. He is preparing a doctoral dissertation in Military Science (leadership). His research interests are information environment, strategic communication, and information operations.

John Hultquist leads the analysis team that tracks cyber espionage threats for iSIGHT Partners' government and commercial clients. Mr. Hultquist has over six years experience in covering cyber espionage and hacktivism, working in senior intelligence analysis positions with the US government prior to iSIGHT Partners.

Margarita Jaitner graduated from Swedish National Defence College with a BA degree in Political Science and an MA in Societal Risk Management from Karlstad University (Sweden). Currently she works at the Yuval Ne'eman Workshop for Science, Technology and Security in Tel Aviv, Israel. Her current research focuses on social media as a toolbox for information warfare.

Tugce Kalkavan is an M.sc. candidate and a Research Assistant at Yasar University, Izmir. She graduated from Computer Engineering from Yasar University at 2013. His research interests are Computer Security, Cryptography, Information Theory, Cyber Warfare and Cyber Espionage.

Harry Kantola has studied (Masters of arts) in Swedish National Defense College (SNDC) and to General Staff Officer at Finnish National Defense University (FNDU) 2011. Currently he is carying out doctoral studies and teaching at the department of tactics and op-erational arts at the same university.

Anthony Keane, MSc, PhD has a background in Astrophysics and received his PhD from work completed at the Dublin Institute for Advanced Studies. He is currently the Head of the Department of Informatics in the Institute of Technology Blanchardstown, and is also the principal investigator of the ITB research group in Information Security & Digital Forensics.

Ilker Kilaz received his B.S.degree in System Engineering from Turkish Military Academy in 2003. He received his MBA in Financial Management from the Naval Postgraduate School in California in 2011. He currently continues his study at the Turkish Army War College. He is interested in leadership & management studies, manpower planning and cyber warfare.

William Aubrey Labuschagne started his career in tertiary education in the fields of programming, networking and security. In 2010 he moved to the CSIR and since then contributed on various projects within Cyber Defence. Currently he is a technology researcher in the fields of social media, social engineering and cyber awareness. He is currently completing his masters on how to improve the effectiveness of security awareness programs.

Brendan Lawless is Chief Executive Officer for the Irish Honeyn3t Chapter, Research Student in the Blanchardstown Institute of Technology Ireland and an advocate for the ethical hacking community. His interests are in the security field where it's fast paced results and consequences keep him always learning and enthusiastic.

Karie Nickson is currently a PhD candidate in Computer Science at the University of Pretoria South Africa focusing on Digital Forensics. Mr. Nickson has presented several research papers at different international conferences and published others in world known scientific journals. His research interests include network security, intrusion detection, information security, and digital forensics.

Kadir Ozyurt received the B.S. degree in System Engineering from Turkish Military Academy in 1997. He received M.S. degree in social and cultural sciences in 2010. He is currently teaching at Department of Operations and Intelligence at Turkish Army War College. He is interested in cyber security and digitization of battlefield.

Kosmas Pipyros is a graduate of the Faculty of Law and Economic Sciences of Aristotle University of Thessaloniki and a graduate of the Hellenic Corps Officers Military Academy of Thessaloniki. He holds an MSc in Information Systems from the Department of Informatics of Athens University of Economics and Business and he is a PhD Candidate of the same department. His current research interest includes the legal and technological dimension of Cyber Warfare. On a professional basis, he has been practicing Law as legal advisor since 2006, specializing on issues of international law in the Armed Forces.

Dirk Ras is currently a Masters student at the University of Pretoria in South Africa. His main area of research is that of digital forensics in cloud computing systems, specifically the monitoring of such systems which influences the forensic readiness of large scale corporate systems.

Vítor J. Sá holds a Computer Science Degree from the University of Minho, where he lectured for several years. He lived for four years in Germany as a guest researcher at the Fraunhofer IGD, in Darmstadt. He completed his PhD work in biometric authentication and, currently, he is a Lecturer at the Catholic University of Portugal.

Henrique Santos received his first degree in Electric and Electronic Engineering, by the University of Coimbra, Portugal, in 1984. In 1996 he got his PhD in Computer Engineering, at the University of the Minho, Portugal. Currently he is an Associate Professor at the University of Minho.

Armin Simma is a senior lecturer at the Vorarlberg University of Applied Sciences, Austria. After graduating at the University of Linz, Austria he worked 2 years at CERN, Switzerland. Since 2001 he is teaching and doing research in the areas of IT security, virtualization, cloud technology, operating systems and computer networks.

Elena Sitnikova is an experienced researcher and academic from the University of South Australia. Prior to joining the university she worked in the IT and Space industry. She has a broad knowledge of software and systems engineering, quality assurance and enterprise process capability. Her current research interests include critical infrastructure protection and SCADA systems cybersecurity.

Aelita Anthimos Skaržauskienė was the couch in the Self-managing team's project in European Parliament together with DEMOS Group Belgium . In her work dr. A. Skaržauskienė applies both knowledge of management and modern leadership-correlated disciplines such as Business dynamics, Systems thinking, Chaos and Complexity theories. Her current position is Director of Business and Media School together with Middlesex University, UK.

Anthimos Alexandros Tsirigotis is a third year Phd student at the University of Reading. His research concerns the changes in the discourse on modern war as a result of the Global Information Age. His area of interest is the United Kingdom and he applies Corpus-Assisted Critical Discourse Analysis.

Hein Venter is an Associate Professor and research group leader at the Information and Computer Security Architectures research group at the University of Pretoria's Department of Computer Science. He holds a PhD in Computer Science from Rand Afrikaans University (now University of Johannesburg). His research interests include network security, intrusion detection, information privacy, and digital forensics.

Stilianos Vidalis, BSc, MSc, PhD is a lecturer at the University of Staffordshire. He received his PhD in Threat Assessment under the context of Information Security in July 2004. His research interests are in the areas of Information Security, Information Operations, Digital Forensics, Threat Assessment, Profiling and effective computer defence mechanisms and he has published a number of papers in those fields.

Suné von Solms obtained her PhD in 2012 from the North West University in Potchefstroom, South Africa. She is a researcher at the Council for Scientific and Industrial Research in Pretoria, South Africa and an Extraordinary Senior Lecturer at the North West University. Her interests lie in applied research in the field of Cyber Security.

Sebastiaan von Solms ('Basie') is a Research Professor in the Academy for Computer Science and Software Engineering at the University of Johannesburg. He is also the Director of the Centre for Cyber Security at the same university. He has published extensively and is the immediate Past President of IFIP, the International Federation for Information Processing.

Murdoch Watney is an admitted advocate of the High Court of South Africa. She contributed to three textbooks and has published extensively nationally and internationally on the law of criminal procedure, criminal law, law of evidence and cyber law in law journals and has delivered a number of papers at national and international conferences.

Cagatay Yucel is a Ph.D. candidate and a Research Assistant at Yasar University, Izmir. He graduated from Computer Engineering from Izmir Institute of Technology at 2009. He got his master's degree in Computer Engineering from Yasar University at 2012. His research interests are Computer Security, Cryptography, Information Theory, Cyber Warfare and Cyber Espionage. He can be contacted at cagatay.yucel@yasar.edu.tr.

Possible Scenarios and Maneuvers for Cyber Operational Area

Ugur Akyazi
Turkish Air Force, Ankara, Turkey
uakyazi@hvkk.tsk.tr

Abstract: Cyberspace is considered to be a military operation area such as land, sea, air and space in most countries. Cyber weapons which are computer codes designed to threaten or harm the systems or human beings physically, functionally or psychologically; can be used either to create kinetic effects as a facilitator for conventional target attacks or cause large-scale death or destruction by modifying the parameters of a system or database. It is not able to occupy the enemy territory only with cyber warfare without a physical combat, but a severe asymmetric damage can be given to the enemy by exploiting the surprise attack or the far distance principle. If both a ballistic missile and a computer virus have the ability to disable a target, the virus will be preferred naturally. The real threat to armed forces is the closed military networks being exposed to cyber attacks via electromagnetic energy. In this study, possible scenarios for cyber operational environment and measures to be taken are examined through doctrinal and operational aspects. The required skills to gain and maintain the superiority over this new area are discussed by evaluating cyber power as a decisive force multiplier when used carefully and correctly at appropriate time.

Keywords: cyber attacks, closed networks, cyber electronic warfare, cyber planning and maneuvering, cyber strategy and doctrines, cyber command

1. Cyber operation environment

Cyber attacks are now not only limited to sneaking into the systems and stealing information, but they also appear as an operation type which has the capability of damaging the communication and computer systems, energy and transportation networks, military command and control systems of a country. Therefore, cyber attacks are one of the most important threats that we will face in the future. Cyber attack incidents that have come up recently, proved that cyber attacks can cause a devastating effect on the systems that are free from internet and closed to the outside world.

Cyber power is the ability to use the cyber space in order to gain advantage by affecting the events on each theater of operations and power elements. As it can also be understood from this definition, now in many countries, cyber space is considered as a military theater of operations like land, sea, air and space. Different from other conflict areas, cyber space provides efficiency on every one of the national power elements. Internet is one of the most important elements that constitute the cyber space; however cyber space is not just limited to internet. Many systems and hardware such as communication networks, military networks close to outside world, energy distribution networks, cell phones, software-based walkie-talkies, electronic command systems, satellite systems and unmanned aerial vehicles are the elements of cyber space.

In this paper, strategic, operational and tactical aspects of the cyber operational area are examined together in order to show the big picture of possible cyber wars. It is tried to answer why cyber defense issues are to be integrated into the Armed Forces structure and planning processes. The cyber concept, doctrine and applications of the nations and organizations (especially USA and NATO), related articles, books and thesis are considered comparatively to reach the suggested recommendations. Both national/international cyber exercises and the real events are closely studied to find out the following case scenarios of the cyber warfare.

1.1 Characteristics of cyber space

In order for cyber security operations to succeed, it is essential to pay attention for the differences between classical and cyber threats, and the own properties of the cyber environment. One of the difficulties is to detect the perpetrators that carry out cyber attacks and the places from where these attacks are organized (Yasar N., 2012). Nowadays, it is not that hard to detect the classical threats and the places where these attacks arise from. For instance; the place, capacity and danger scale of a biological, chemical and nuclear weapon facility can be detected. However, because the perpetrators of cyber attacks are using anonymizing programs that make it possible for hiding themselves in the electronic environment, they are hardly detectable. Due to the existence of mobile communication technologies and their ongoing development, and because it is possible to carry out attacks using computers such as botnets (slave computer networks) whose

control are captured; these attacks are not based on any place and they also make the concept of attack center meaningless.

Another point is the access power and range of cyber attacks. Classical attack tools do have a certain capacity and these tools are able to pose a threat within the range. For instance; a missile can only pose a threat for the cities, countries and regions within its range. In order to enhance the effects of classical attack tools such as a missile, advanced level information and experience along with economical power corresponding to the level of billion dollars are required. However, nowadays cyber attack tools can be enhanced with a few amount of information and money; and cyber attacks can be organized by using these tools from east to west, north to south of the world via internet. In this respect, discrimination in the form of national and international dimension is not such meaningful in terms of cyber security (Unver M., 2009).

Cyber space definition of USA Armed Forces (U.S. Department of Defense, 2011) shows that cyber space includes much more than computer networks. More importantly, as network systems are created using electromagnetic energy, wars in cyber space will also include the skills that are currently regarded as electronic warfare. By using electromagnetic energy, it would be possible to access and attack the networks which are not connected to the internet directly, in order to query and break the electronic components of a network. This is a serious threat for armed forces.

1.2 Cyber war

Cyber war is not alike conventional war; however it shares several characteristics such as air bombing, submarine battle, special operation forces and even the historical role of assassins. Particularly, it may cause asymmetric harm to an enemy critically from a far distance or by exploiting the surprise principle. It is not possible for cyber attacks, at least in the near future, to be more lethal than a strategic bombing. But ultimately, the success of military operations is effect-based. If both ballistic missile and a computer worm can destroy or deactivate a target, the natural selection would be the computer worm (Geers K., 2011).

Without a physical intervention, it is not possible to invade the enemy territory with cyber war only. Because one cannot take hold of the other's weapon, the fight takes place and goes on like "who is going to shoot longer". At this point, in order to be effective in a conventional war, carrying out cyber wars along with the conventional ones comes into prominence. Martin Libicki classifies the cyber war as strategic and operational (Libicki M.C., 2009). Strategic cyber war is the cyber attack campaign which is directed towards changing the attitude of a government. The assumption here is that, there is no other hostile war going on between two sides except for the cyber war. Otherwise, if cyber war is carried out in a way that supports the military operation, it becomes operational cyber war. Cyber attacks are not capable of causing great damage to the other side, when compared the nuclear and strategic conventional weapons. However, even if operational cyber war does not constitute the power itself, it could be a meaningful power factor if applied carefully, cautiously and correctly in appropriate time (Cifci H., 2013).

One of the most important elements in the cyber war is intelligence. It is necessary to collect detailed information about the properties, connections, hardware and software content of the target system and to reveal the security deficits in these elements. Making a delicate assessment about whether the cyber attacks are effective or not, is another problematic field in the cyber war. It is quite difficult to detect which systems are affected in what extent, or even whether they are affected or not. Knowing how the opposite systems operate after the attack is essential to understand whether the system is broken or not.

In order to utilize the skills of Air Forces for the purpose of carrying out operations in the cyber space, USA Air Forces Cyber Command was established. Air Forces aims at directing, commanding and controlling the enemy forces in a joint war, and even making it possible to create actual effects in cyber space in a way that would affect the firing capability of their weapons. These are not virtual or imaginary effects – the enemy might really die or get hurt in a warfare that takes place in cyber space (Fahrenkrug D.T., 2007).

R. Clarke claims that 600 hackers are working under the command of Lab 110 unit in North Korea Armed Forces Cyberwar Command. There are 100 hackers working in unit no. 204, which is secret cyber psychological war unit. The mission of unit no. 121, whose connections are in China, is to take down the military command, control and communication networks of South Korea. As internet in North Korea is too limited and because it

is easy to catch, the activities are conducted via China. Members of these units are employed starting from the level of primary school. These students enter programming and hardware departments in their further secondary education and then, they are accepted for the Automation University. The only focus of this university is hacking the network systems of the enemy (Clarke R, 2013).

After defining the cyber space and telling about some elements of cyber warfare, the conceptual and legal parts of these notions are discussed in the next chapter.

2. Conceptual and legal thoughts about cyber warfare

In USA Air Forces Basic Doctrine (AFDD-1, 2011), air power is defined as "the ability of leading the military power or effect by means of using and controlling air, space and cyber space in order to reach strategic, operational and tactical targets". Cyber Space Superiority which is one of the basic functions of the Air Forces is stated as "operation advantage in the cyber space for the purpose of carrying out operation at a particular time and domain without a blocking intervention (AFDD 3-12, 2011)". Sub-entries of this function are: Using Cyber Space Forces, Cyber Space Defense and Cyber Space Support.

USA Secretary of Defense officially declared that cyber space is conceived as a war zone by publishing the National Military Strategy for Cyber Space Operations. In this document (U.S. Department of Defense, 2011), cyber space is defined as a domain in which electronic and electromagnetic spectrum is used in order to hide, change and provide data stream thanks to the network systems and relevant physical infrastructures. Cyber space is among other domains which are air, land, sea and space; and it connects the hidden, changed or transmitted data with these physical domains.

The decision of using armed forces to carry out cyber attacks is subjected to at least three conditions; perpetrator of the attack (target definition), legal dimension (preparing the legal conditions and requirements of the attack) and skills (determining the instruments and methods required to carry out the operation). The issue of "Perpetrator of the attack" is being discussed for a couple of years. The view that it is impossible for carrying out operation on cyber attackers or the places where cyber attack takes place is widespread. In addition, many professionals have recently stated that the situation is different in fact. Dmitri Alperovitch, former vice president of Threat Research Unit of McAfee Corporation which is the provider of contracted standard information systems security of USA Secretary of Defense, has stated that it is not a problem to detect the perpetrator of an attack. He also added that "There is large-scale intelligence information available that appears in the context of informatics crime investigation (forensics) which makes it easy to determine the perpetrator of an attack.

However, is it legal to counteract by only depending on this foundation? If so, in which level it is proper to counteract? Who should counteract? And according to which laws? In which level [cyber] threat becomes a national security threat and requires the Secretary of Defense to do what it always does? As a reply to some of these questions, The Tallinn Manual on International Law Applicable to Cyber Warfare – MILCW was prepared by NATO Cooperative Cyber Defense Centre of Excellence – CCDCOE (The Tallinn Manual, 2013). According to this manual, a cyber operation which is controlled by a government and concerning the infrastructure settled in another country might violate the sovereignty of that country; in state of damage, it definitely does. If these types of operations have the intention of pressuring the government, they might constitute the use of prohibited "intervention" or prohibited "force usage". A cyber operation which is called as an armed attack triggers the use of individual or total self-defense rights.

Next two chapters discuss the cases and scenarios of this new war, and suggest to plan and maneuver like the air operations which would bring the victory.

3. Possible scenarios concerning the operation environment

In a report published in Defense News magazine dated 15 January 2013, it was mentioned that US Army Intelligence and Information Warfare Directorate was making the study about the penetration into closed networks via electromagnetic waves. It was stated that Tactical Electromagnetic Cyber Warfare Demonstrator program which is conducted in this context had two purpose; one to steal data via closed cable networks, and the other to leak data into these networks. This technology would be possible by means of a drone flying over the area where the target networks are located. Despite the successful results acquired in the tests carried

out, this technology is not yet ready for use due to the limits of efficient range and bandwidth. The source of waves must be very close to the target network and the transfer of complex data takes a long time.

Especially in terms of understanding the point better, Electronic Attack (EA) within classical EW and Cyber Electronic Attack should be compared. During classical EA, we jam the instrument of threat that sees us. For instance, radar of a SAM system is targeted. Therefore, the threat could be jammed in the time frame when it looks at the platform where the threat is tried to be protected. Thus, whenever the threat system looks at the protected platform, the jamming signal needs to be reproduced with the same efficiency, and transmitted continuously. However, in a Cyber EA the target would be the central processor unit of that system, in other words 'its brain'. The real purpose here is to send a software code to the threat system's brain, as a result of the attack to be carried out. Different from classical EA, it would be sufficient to accomplish this process only once. Another point which should also be stated is that; when Classical EA is applied, the threat would notice that it was exposed to jamming and would be able to take necessary precautions; however, when a Cyber EA is applied, it would not be noticed that easily or it would be too late when it is noticed.

A similar goal is pursued in 'Suter' project, run by BAE Systems. "Suter" project which was managed by 'Big Safari', a secret unit in US Air Forces, was integrated in UAVs and had been tested in Iraq and Afghanistan since 2006. It is considered that Israel also used a similar technology in order to blind the radars when bombing Syrian nuclear facility in 2007.

Unmanned aerial vehicles (drones) that are developed for military and intelligence purposes seem to face difficulty when stepping in the public domain. After a short while when the Amazon Corporation announced its project about transporting its cargo with drones, hacker Samy Kamkar introduced an invention, serving as a warning. The drone called Skyjack which is developed basing on the model Parriot AR Drone 2, brings the unmanned aerial vehicles under control by means of the wireless signals transmitted during flight. Kamkar, who made his name with the hacker attack that had taken down MySpace in 2005, and then detected a lot of security gaps in many fields including credit cards, run SkyJack with the smallest computer of the world, Raspberry Pi. The program that he has loaded into the aircraft, in which he uses USB battery and two wireless adaptors is Linux based and it has been developed to deviate the drones.

Cyber attack of SkyJack that inebriates the drones has a simple process. First of all, one of the wireless adaptors detects nearby wireless connections and disconnects the unmanned aerial vehicle that enters the network coverage. Then, the other adaptor connects the idle drone with itself via the network it has created. Later on, the 'zombie drone' can be directed as desired by Javascript commands sent from tablet computer or laptop. This system can be applied on the mobile platform and on the desk subjecting to the network coverage (Kamkar S., 2013).

Another serious threat is the authorized users in the network who have the advantage over the outside attackers in terms of damaging the corporation. Fire walls, intrusion detection systems and access control systems are generally executed against the threats that would come from outside. However, network users might be informed about the information security policy, methods and technology of the corporation as well as the gaps inside these systems which could be exploited.

By the "Cyber Insider Threat-CINDER" program being developed by US Defense Advanced Research Projects Agency-DARPA Information Innovation Office, it is aimed at determining the cyber espionage operations that might be conducted in the military networks. Raytheon Company of USA argues that an effective solution of a cyber threat should be one step ahead on the threats proactively. "SureView' program developed for this purpose, not only focuses on the behaviors of malicious codes in the network, but also the security policy breaches of the users. Collected user data such as keyboard typing, mouse movements, opening documents or visited web sites can be watched as video in almost real time. By keeping the fingerprints of the corporation and the critical files of the network, this program can also warn the intrusion attempts to be made.

It should be considered that below scenarios concerning the operation environment might take place:

- Changing digitally the coordinates of the secret level air pictures,
- The misinformation about the operation being published on the official page of the Joint Chief of Staff, after DNS (Domain Name Server) servers had been captured,

- Making the drone systems get out of the control or capturing the expedition images acquired after a virus attack in the ground control station of these systems,
- Taking over the control of satellites, preventing and directing their communication systems as a result of the cyber attack in the ground control systems of the satellites,
- Changing, stopping or erasing the data stream after Tactical Data Link systems were jumbled when their crypto and codes were deciphered,
- Affecting the operations with special viruses produced against electricity generation and distribution systems,
- Giving false information to the aircrafts that would land on the square, and causing physically great damages to them by sending cyber attacks to the computers of navigation systems such as GCA, RAPCON, TACAN and INS.

4. Planning and cyber maneuvers

Logical maneuver capability in the cyber space is a function of security protocols that are generally used by the host computer systems. Defense against the entry of unwanted systems is included in the code or logic of the host computer's system. After a connection is established among the systems, a potential attack needs to utilize a logical error in order to enter the system. Therefore, logical maneuver format in the cyber space is code-writing (Akyazi U., 2013). The attacker writes malicious codes in order to acquire maneuver capability against the targeted systems. When an unwanted entity is noticed in the system, the security system would change the system code to prevent the entry. The attacker, desiring to remain in the system, customizes the malicious code in this direction. This process is the equivalent of the forces that maneuver in order to acquire advantageous positions in the fields of traditional air, land, space and marine theatres. In Table-1, examples of the applications of Air and Space Power Usage Principles in the Cyber Space Operation are given.

Planning and execution of air operations are carried out by CAOCs which are the decision and control centers in operative level. Having the decision superiority that has critical importance for the success of the operation in CAOC where 1000-1500 flights are planned and their execution are followed daily during the operation, appears to be a force multiplier. The purpose of operation is a decision that is made depending on political evaluations and that directs the operation planning process. Planning process executed by CAOC starts in this process. Following three stages; target selection, ammunition choice and aircraft allocation can be regarded as the most important stages for the success of the operation. Selected targets are needed to provide the maximum contribution to the final purpose of the operation, using the most convenient sources within the shortest time. After the targets are selected, suitable ammunition to be used will be allocated. At the next stage, the allocation of bomber aircrafts that would carry the relevant ammunition, damage the enemy targets and contribute directly to the purpose of the operation, will be carried out. Finally, Air Task Order that includes assignment details of the aircraft and ammunition to be used for attacking the selected targets will be produced, and then execution begins (Arslan O., 2012).

Air Forces Cyber Task Order (CTO), a significant product of the duty cycle, is used by the cyber space forces to receive and carry out their missions. Cyber duty cycle depending on air task cycle is an iterative process which is used for the planning, coordination, execution and evaluation of cyber space operations. In order to support the crises, duration of the cycle could be extended or shortened in accordance with the course of the warfare. In order to prevent the conflicts and preserve situational awareness, the whole cyber operations should be involved in CTO. At the planning stage, in order to complete the CTO, Operation Center uses COMAFFOR / JFACC directive, rules of engagement (ROE), joint priority target list.

Legal issues and international legal liabilities are also applied in the use of cyber capabilities. International and municipal law, political decisions, law of armed conflicts and rules of engagement do constitute the legal framework in which the operation activities are evaluated. Generally, the military legal expert in the quarters of the commander, counsels the relevant commander about carrying out cyber space operations legally. Legal support personnel should have access to information, process and programs that are used in cyber space operations, and also mastery on the basic cyber technologies (AFDD 3-12, 2011).

Table-1: Cyber operation applications of air and space power usage principles

PROPERTIES	PURPOSE	SAMPLE CYBER SPACE OPERATION
Central Control, Decentralized Execution	Control by an air commander who has a broad point of view; execution by people who best conceive the tactical specifics of a dynamic operation.	Cyber operation concept of the commander changes into actions that are executed by Regional Combined Air Operation Centers (CAOCs) and local Network Operation Centers.
Flexibility and Versatility	Applying the weight center and maneuver principle in each level of the warfare at the same time.	Flexibility and versatility are included in the own nature of cyber space. A small piece of code could create tactical, operative or strategic impacts, depending on the target.
Synergistic Effect	Combining the use of forces in order to create more impacts from the individual contributions of power elements.	Supporting the solid, continuous and sustainable connections along with the integration of C4ISR that support real-time joint operations.
Continuity	Executing continuous operation that enables shooting the targets over and over again, in a level that is close to desired.	Distributed Denial of Service Attacks (Going on until a conscious respond is given)
Concentration	Focusing the penetrating power at the right time and place.	Simultaneous cyber attacks at more than one network
Priority	Determining the right priorities for the use of air power.	Prior ISR requirements
Balance	Balancing the opportunity, necessity, effectiveness and productivity against the risks owned by friendly forces	Air Force Networks can be used by the staff of Air Force and other forces, non-governmental organizations and military members of coalition (if required) in order to meet the national security requirements

5. Results and recommendations

Cyber space is expected to have effects on the operation environment in each level. Correspondence between the troops of Air Forces Command and the operation center is provided via a secure communication network. Air operation can be planned via operation, intelligence and flight training modules included in the Information System that uses this communication network. During the operation carried out, information related to the surface and air components can be followed closely to the real time, and this information can be transferred to the Information System via air defense system by means of land, marine and air platforms using Link-1, Link-11 and Link-16. Thus, the entire operation area is exposed in the Information System and situational awareness is provided. If a system that provided such a great integration is exposed to a cyber attack, the use of air power would be disrupted in a great scale. Therefore, it might also affect the use of air power in the land and marine operation. Thus; it is considered that Air Forces might be affected from a cyber attack more than others and it is necessary to determine the cyber threats which concern the air power in detail, as Air Forces especially uses the modern technology and Information System actively.

Most of the operating system, application software and hardware that are both used in open and closed networks are composed of commercial products. No matter how their system entry/exit opportunities such as USB, CD and printer/scanner are limited and controlled, data entry/exit permission is given to the computers in closed networks. Using the same CDs also in the internet might cause the closed network to be exposed to the attacks in the internet, because malicious software can hide themselves inside the files and infect into the systems via these programs. Systems that have the capability of stealing data from closed cable networks and leaking data into these networks are being developed by means of an unmanned aerial vehicle that flied over the area where the target networks are located. Therefore, all the security software in closed networks should be installed and updated.

As it is included in NATO Cyber Defense Policy (2011), cyber defense issues must be integrated into the Armed Forces structure and planning processes; awareness programs should be developed and cyber components should be added to all military exercises. For the planning, coordination, execution and evaluation of cyber space operations, a cyber duty cycle depending on air task cycle could be used. Cyber security cells should be added to the operation centers and cyber operation centers that support the operation should be established. In addition to the technical elements such as information systems security, communication network security and forensic analysis; executive or consultant personnel that would serve on the subjects of intelligence, operation and law should be assigned. As knowledge and experience is gaining importance in cyber defense, a sub-domain of "Cyber Defense / Information System Security" should be established within the armed forces in order to provide the sustainability of expert and qualified personnel on the subject.

References

AFDD-1 (2011) Air Force Basic Doctrine: Air Force Doctrine Document 1, USA.

AFDD 3-12 (2011) Cyberspace Operations: Air Force Doctrine Document 3-12, USA.

Akyazi, U. (2013) "Preventive Measures in the Context of International Cyber Security Strategies and Doctrines", Paper read at 6th International Information Security & Cryptology Conference, Ankara, Turkey, 20-21 September.

Arslan, O. and Yucel, E. (2012) Air Task Order Optimization Model, Int. Journal of Aerospace Technology, Vol 5, No 3, January, Istanbul, Turkey.

Cifci, H. (2013) Her Yönüyle Siber Savaş, Tubitak Press, Istanbul, Turkey.

Clarke, R. and Knake, R. (2012) Cyber War: The Next Threat to National Security and What to Do About It, Ecco Press.

Fahrenkrug, D.T. (2007) "Cyberspace Defined", The Wright Stuff, Air University, 17 May.

Geers, K., (2011) "Strategic Cyber Security", NATO Cooperative Cyber Defence Centre of Excellence

National Cyber Security Framework Manual, Talinn, NATO CCDCOE Publication.

Kamkar, S. (2013) "Skyjack: Autonomous Drone Hacking", [online], http://samy.pl/skyjack

Libicki, M.C. (2009) Cyberdeterrence and Cyberwar, RAND Corporation, USA.

Unver M. and Canbay C. (2009) "Cyber Security in National and International Dimensions", [online], www.btk.gov.tr/bilgi_teknolojileri/siber_guvenlik/dokumanlar/siber_guvenlik.pdf

U.S. Department of Defense (2011) Strategy for Operating in Cyberspace.

Yasar N., Yasar F.M. and Topcu Y. (2012) "Operational Advantages of Cyber Electronic Warfare (CEW) in the Battlefield", Paper read at Cyber Sensing 2012, SPIE, Baltimore, Maryland, USA

Digital Forensics as a Science in Higher Education

Olga Angelopoulou[1] and Stilianos Vidalis[2]
[1]Faculty of Business, Computing and Law, University of Derby, Derby, UK
[2]Faculty of Computing, Engineering and Sciences, Staffordshire University, Stafford, UK
o.angelopoulou@derby.ac.uk
stilianos.vidalis@staffs.ac.uk

Abstract: Digital Forensics is a multi-disciplinary applied science 'governed' by legislation. From an academic perspective the nature of digital forensics creates a number of challenges in the classroom. It requires the mentor-tutor to include criminology, psychology, sociology, law, forensic science, mathematics, and computer science in his expertise and develop and maintain an airtight learning environment that promotes freedom and academic excellence while ensuring and assuring its external boundaries (in a continuous manner) by employing appropriate forensic controls. This introduces a difficulty and complexity in the mentoring capabilities of the academic that is seldom recognized by management. In our paper we discuss our lean approach in educating digital forensic investigators, present our learning environment and argue that playing games is indeed the best method for transferring skills and knowledge to the different types of scholars studying towards an academically accredited digital forensics qualification. We also attempt to identify whether professional qualifications in the field of digital forensics could enhance the student experience and improve the student employability.

Keywords: digital forensics, digital investigation, higher education, learning environment

1. The challenges

Digital forensics is an amalgamation of forensic science and computer science and is governed by strict and rigorous rules and regulations. According to (Schweitzer, 2003) digital forensics is

> *"the science of acquiring, retrieving, preserving, and presenting data that has been processed electronically and stored on computer media."*

Palmer (2001) defines digital forensics as:

> *"the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations."*

Individuals pursuing a career in this discipline are required to have an 'enriched' background drawing elements of practical experience from fields as varied as sociology, psychology, forensic science, computing and the law. The challenge faced by educators is to train individuals, many of whom have no prior theoretical or practical experience in the aforementioned fields, in the following skills:

- Sound understanding of the forensic science principles
- Practical knowledge of the different types of computer crime
- Understanding and practical knowledge of managing a digital crime scene
- Sound understanding and practical knowledge of managing evidence of user activities
- Ability to use with competency digital forensic toolkits
- Ability to follow strict policies and procedures with meticulous record keeping
- Good understanding of people and their motivational catalysts.
- Knowledge of evidence law and legal procedures.
- Ability to write reports on technical issues in a non-technical manner
- Ability to address large audiences in a formal manner and affect their decision making process.

Plato in *the Republic* indicates that there are two different forms of vision, the *mind's eye* and the *bodily eye*. The *bodily eye* is a metaphor for the senses of the prisoners in the cave. We use the *prisoners* as a metaphor for our full time students that operate in the protected academic environment of their University, which is the

cave. The *mind's eye* is a higher level of thinking, and is utilised only when the prisoner (the student) is released into the outside world and in this case when the student graduates and gets a full time job. When students graduate they are forced to act in a 'real' corporate environment and combine their knowledge to solve real life problems; that's when the students can truly learn, develop and use their higher level of thinking. Hence, the first challenge for the mentor-tutor is to encourage and enable the students to use their *mind's eye* during their University studies.

In order to address the aforementioned challenges, and the learner's requirements discussed at a later section, we have developed an educational approach that is experience based (Kolb (1984), Boud et al. (1985)), learner centred (Tudor, 1993) and problem based (Schmidt, 1993) for enabling the students to understand and appreciate the inter-disciplinary nature of the digital forensics, and be able to link the theories with the real problems in the field.

We also argue that for students to learn effectively and in depth, they need to feel as being valued and belonging. It is only then that they are likely to be able to engage with the business of learning. Hence the second challenge for the mentor-tutor is to establish a certain culture, following the strict principles that are discussed later, and the third challenge is to actually promote that culture amongst the student cohorts.

This work presents our approach to the delivery of the digital forensics programme to the students. We also investigate areas that a professional certification could be incorporated in higher education and how this could be implemented. The principle is the current curriculum practise in digital forensics offered at our institutions. The study is lead mainly by observing the student achievement throughout our experience in academia. We explore the current teaching and learning methods and a combination of methods that include elements from the professional qualification courses.

Therefore, we review the current literature in learning and teaching in relation to digital forensics and the professional qualifications that are offered, we attempt to identify the academic and career benefit of professional qualifications in the area of digital forensics for students in Higher Education and discuss the curriculum implications towards the current practise.

2. The learning axioms and models

The literature shows three schools of thought in learning and teaching. The Behavioural approach examines how the student behaves and the focus is to create a good environment that can facilitate teaching and learning based on individual needs. The Humanistic approach encourages *self-knowledge* and *personal development*. The Cognitive school of thought pays attention on *how knowledge is acquired* in order to contribute to student learning. (Cruickshank et al., 2009)

Using the principles from the aforementioned schools of thought, and data collected from delivering undergraduate and postgraduate modules in the field of digital forensics since 2001 we have concluded to a number of learning axioms:

- Learners need to understand the relevance of what they are learning and how it relates to and has relevance to their future employment.
- Learners prefer to take responsibility for their decisions and actions and value the ability to self-direct their learning.
- Learners have accumulated a great volume of experience over time, which could enhance their learning and necessitates individualisation of learning strategies.
- Learners relate better to things that they can equate with real-life problems.
- Learners have a task-centred orientation to learning and like to feel free to focus on the task or problem.

Clearly, we refer to learners studying towards an academically accredited digital forensics qualification offered at a University in the United Kingdom. Each school of thought has inspired academics in developing different learning models.

Bloom's taxonomy of learning (Bloom and Krathwohl, 1956) suggests a model that classifies thinking according to the level of complexity required by the learner, the hierarchy of the cognitive domain to an individual's

learning. His taxonomy is still valid after all these years due to the fact that it clearly sets the goals and objectives for creating learning outcomes. His model has a hierarchical approach, where the students need to 'master' a base level and continue to the next one. It is considered as a traditional teaching and learning being satisfied by six objective levels. The knowledge, comprehension and application of the subject area are the lower levels of the model, where the learner respectively remembers, understands and applies the knowledge. The three upper levels of his hierarchy are the analysis, evaluation and creation, aiming to promote the critical thinking of the learner (Ennis, 1993). The taxonomy as demonstrated in figure 1, is considered as an effective model that 'has given rise to the educational context' (Forehand, 2005).

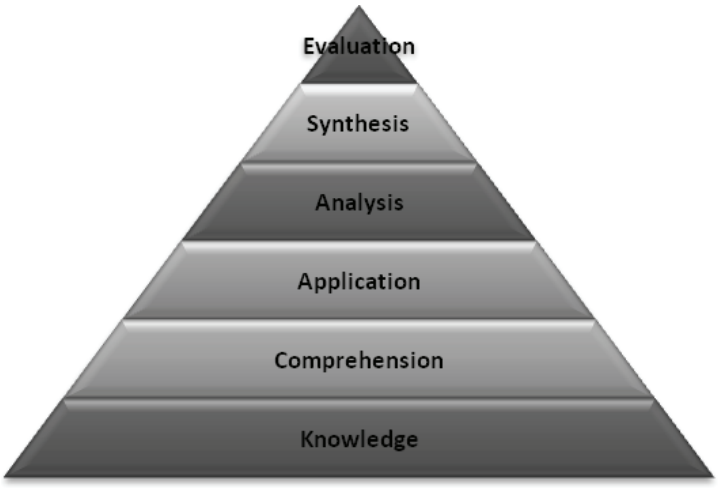

Figure 1: Bloom's taxonomy of learning

Bloom's taxonomy follows the cognitive school of thought, and the aim was to influence the cognitive domain. However, later versions of the taxonomy discuss skills in the affective (Krathwohl et al., 1964) and psychomotor domain (Simpson (1972), Harrow (1972)). The aim of the additional work was to measure the knowledge, the attitude and the skills of the learner. It can assist in building the learner centred and the problem based approach by constructing the appropriate learning outcomes for the students.

Kolb (1984) supports the behavioural school of thought and effectively mentions that there are a number of learning and teaching views and styles, because different types work for different people. His model is flexible and adjusts to the needs of his cohort, see figure 2. He supports the experiential learning and claims that

> *"learning is the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping experience and transforming it."*

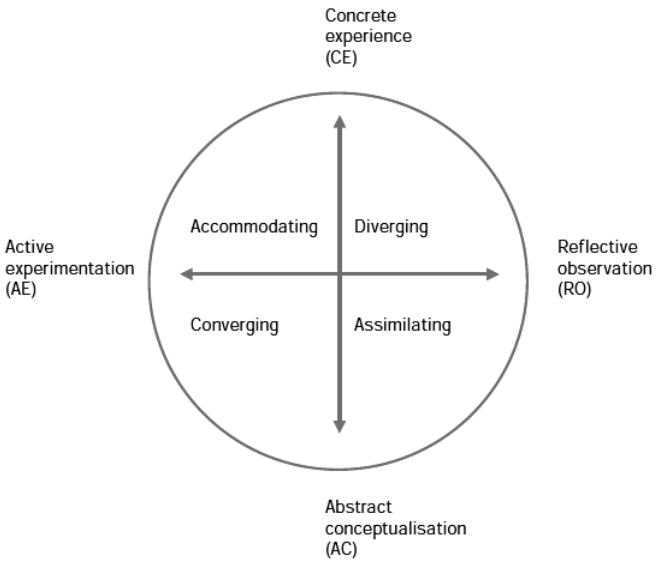

Figure 2: Kolb's learning styles (Coffield et al, 2004)

He parallelises learning with a circle in four stages that lead the learner to four forms of knowledge:

> *"divergence, achieved by reliance on apprehension transformed by intention; assimilation, achieved by comprehension transformed by intention; convergence, achieved through extensive transformation of comprehension; and accommodation, achieved through extensive transformation of apprehension."*

Kolb's forms of knowledge also define his learning styles (Kolbs, 2000). Under these terms Kolb's learning style can be a tool for experience, learner centred and problem based learning.

Biggs and Collis (1982) presented the SOLO (Structure of the Observed Learning Outcome) taxonomy. A few years later they defined it as 'the level of abstraction that a learner uses when handling the elements of a task (Biggs and Collis, 1989). They tried to set the standards for lifelong learning. The model suggests a continuously developing process for the learner. It extends from the knowledge the learner has not yet obtained to the knowledge he has mastered. Two phases and five stages constitute the SOLO taxonomy. Similarly Bloom's taxonomy, the SOLO can also be represented in a pyramid as in figure 3.

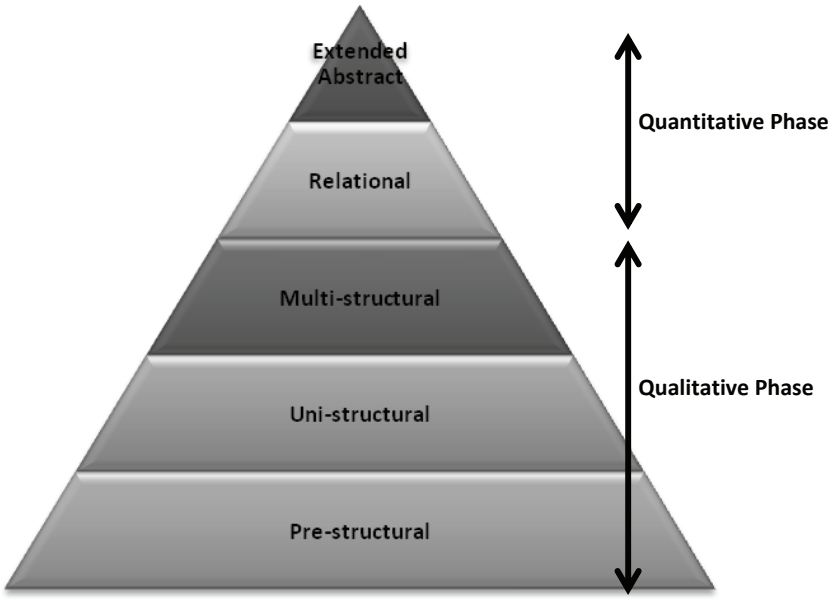

Figure 3: The SOLO taxonomy

On the pre-structural stage students hear, 'but academically show little evidence of relevant learning' (Biggs and Tang, 2007). On the uni-structural phase could be described as the surface learning when the students familiarise with the terminology and describe some points. On the multi-structural phase connections to the subject area have started developing at the same time with ideas. The qualitative phase represents the understanding of the learner. On the relational phase the appreciation of the subject area has developed and ideas can be linked, while on the extended abstract phase the student can link his ideas with the subject area and combine different themes.

The SOLO taxonomy describes how the way of thinking can change through learning. It could be metaphorically interpreted as the development of knowledge from birth to adulthood. In higher education it can apply as the concept of moving from teaching to learning. The cognitive nature of the SOLO taxonomy is also a tool that can be used to achieve experiential, learner centred and problem based learning.

3. Learning digital forensics – our approach

Following Bloom's taxonomy of learning and in par with the British Forensic Science Society we have concluded to the following learning boundaries between each programme level:

- Level 4: The scholar must develop theoretical and practical knowledge of the computer science and of the forensic science principles that will allow for the effective and appropriate management of digital evidence while comprehending the issues that modern technologies introduce in their presentation.

- Level 5: The scholar must apply the theoretical and practical knowledge developed in level 4 for analysing digital evidence regardless of the medium they were collected from, while comprehending the legal issues that relate to the different computing technologies
- Level 6: The scholar must analyse new trends in computing and forensics for synthesising an opinion that will allow for the evaluation of case work.

A waterfall learning approach though is not considered to be appropriate in the current knowledge based computing era, nor it is deemed to be lean (given the hindrances that the modern operational University framework introduces to the learning cycle). A combination of the discussed frameworks, models and theories appears to be the most effective for engaging the students in learning. Based on the behavioural school of thought every student has different needs, but based on the cognitive school of thought we need to assist in the knowledge acquisition and based on the humanistic school of thought we also need to encourage the personal development of each student. To address the above needs we follow a spiral approach (overlaying Kolb's learning styles over Bloom's taxonomy) over the three year duration of the undergraduate degree in using the following learning modes:

- *Reflective learning:* The reflective practitioner theory (Schön, 1991) emphasises relations between knowledge and experience and suggests that individuals reflect in an attempt to link theory to practice. Scholars are reflecting on their activities and processes of working together through team contracts and individual reflection reports that they have to prepare and submit for their formal assessments. The majority of the in-course assessment is group work that reflects over past work that other scholars have completed. Each scholar is keeping a template log book over the duration of the programme that can then be used as a point of reference when conducting digital forensic investigations.
- *Problem based learning:* Biggs (2007) comments that reflective learning has the same aim with problem based learning. Scholars work in small groups examining service user-centred scenarios or case studies that are driven by the needs/problems of our industrial partners. The majority of the in-course assessment is group work that builds over past problems that other scholars have encountered/introduced. The idea is that scholars work cooperatively in order to achieve both subject learning outcomes and transferable teamwork skills as a method of aligned teaching (Biggs, 2007).
- *Experience based learning:* Scholars role-play with other students, guest lecturers and practitioners from the industry as an important part of gaining practical experience throughout their studies. They participate in a three year simulation that is making use of the learning environment that emulates a digital forensics laboratory following industry standards and current practice from the field. Simulation can provide a relatively safe context in which scholars are able to practice skills and receive feedback in a way that would not be possible in a real environment. The pinnacle of this method of learning is the mock trials where scholars present their investigation findings to a magistrate and they are cross-examined by their peers.

3.1 Professional qualifications and curriculum implications

The new higher education era drives us to a competitive and demanding future. Therefore, we need to perceive whether we need to demonstrate quality in the field or superiority amongst our competitors. Irons and Konstadopoulou (2006) discuss about professionalism in the field. There are a number of technical expectations and requirements that need to be considered in a digital forensics course curriculum, but yet the academic perspective should be preserved.

The academics involved in teaching digital forensics need to ensure the delivery of digital forensics address the following potential curriculum implications, such as the practical application of the science, the right balance between theory and practice, and the legal and ethical requirements that need to be considered for a digital forensic investigator. There are also pedagogic requirements, research on the specialist subject and also specific resource requirements.

In order to meet these, there is a need for further research and development in learning and teaching techniques of some disciplines. In specific, the investment in the digital forensics research will enhance the quality and the content of the material produced and offered to the students. This will then lead towards more practical programmes enhanced by professional certifications and accreditations.

In an attempt to meet the requirements for a complete and current academic programme, the curriculum design should pursue the academic value. The Higher Learning Commission (2012) defines the criteria for the accreditation of an academic programme. These criteria could be implemented in any programme regardless the discipline; their core components can adopt in any field, but also outline a complete framework of the values required in digital forensics. Such a framework could also probably adopt certifications and professional accreditations, due to the flexibility of its criteria design.

4. Assessment in digital forensics

4.1 The theory

"Students can escape bad teaching; they can't escape bad assessment" (Boud (1995) in Race & Brown (2001)). The purpose of assessment ideally should be to motivate the scholars, stimulate their interest, allow them and their mentor/tutor to track their progress and provide feedback. However, "*assessment defines what students define as important , how they spend their time and how they come to see themselves as students and then as graduates*" (Brown, 2001), which requires strategic decisions from the academic in setting appropriate learning outcomes and assessment methods. It should evaluate the students' contribution in the area of studies and show them the way for improvement.

Summative assessment is the 'formal' assessment (Fry et al. (1999), Biggs (2007)), guided by the need to measure achievement in relation to the aims and objectives specified for the programme, and the specified learning outcomes of the individual modules. The summative assessment evaluates the scholar's achievement after the learning process.

Formative assessment or 'feedback' (Fry et al. (1999), Biggs (2007)) is the 'informal' assessment, guided by the need to provide scholars with developmental feedback. Of particular importance is staged feedback to enhance performance. This type of formative assessment encourages reflective practice, develops academic and personal skills, builds confidence and promotes deep learning. The summative assessment must be continuous throughout the learning experience.

4.2 The application on digital forensics

Overall, scholar achievement is assessed across the whole programme both formatively and summatively. It is recognised however that achievement of some components may be more difficult to assess, because of the multidisciplinary nature of digital forensics. Nevertheless, graduates should have achieved the 'Extended Abstract' understanding of the subject area through their assessments. (Fry et al., 1999)

In digital forensics it is important that scholars acquire the investigative and goal oriented way of thinking and the 'expert witness' code of conduct. Group assignments and group work will promote team work and will assist the scholars in developing the necessary skills, which relates to the real world and promotes their employability. Scholars can undertake large scale group projects on the full investigation of digital media and produce project contracts, project initiation documents, project plans and personal development planning. This type of assessment is both formative and summative.

Ennis (1993) discusses the value of critical thinking through assessment. In digital forensics critical thinking can be developed through a number of discussion sessions using real life case studies from various security incidents and network attacks. The scholars can advance their incident response skills by analysing the case studies, which can promote critical thinking and critical evaluation. At the same time they can be assessed from their peers and present their research findings. These can take place during tutorial sessions and assess them formatively.

A professional digital forensic investigator will need at some point to present evidence in court, either as an expert witness or a technical witness. Therefore, presentation skills are a prerequisite for the professional. Legal and technical modules aim to cover this need in the terms of content. However, oral presentations of findings are essential to assess the scholar and prepare him for the future. Oral presentations can be recorded (Biggs, 2007) and feedback could be based on those recordings by the tutors and by the student's peers. This technique assists in the development of the communication and the presentation skills of the student. Van den Berg et al. (2006) argue that peer assessment 'resembles professional practice', an asset for the digital

forensics students. Even though oral presentations are considered as a form of summative assessment, the formative peer review can improve this process (Wright and Jones, 2004) and combine both assessment methods.

Assignments as well as portfolios promote deeper learning (Biggs, 2003). Report writing is a critical component for the forensic investigator's work, thus the development of this skill makes it an essential type of assessment on a digital forensics programme. Even though the assignment can provide formative feedback to the student, it is mainly treated as a summative assessment from the student (Covic and Jones, 2007). In the digital forensics discipline specific tasks can be assigned to the scholars that involve independent study, research and critical evaluation of terms and issues in order to promote thinking, analytical and report writing skills.

The portfolio is a selection of work and requires effort both for the scholar to create it and the academic to assess it. However, it promotes creativity (Biggs, 2007) and can also assist in the achievement of the qualitative phase of the SOLO taxonomy. The portfolio is mainly summative assessment; however it can be formative if the academic acts as the facilitator and provides feedback, revises and collaborates with the students (Elbow, 1994). A good example of the portfolio assessment could be a digital investigation of a hard drive, as a practical assignment during practical sessions that is then reflected in a portfolio as a collection of work. The scholars could select their most important findings and include them in their portfolio.

5. An example on the current practise

There has been a lot of discussion at our institutions concerning the official adoption of a professional training as part of the undergraduate programme. There are multiple ways this could be implemented and a number of options have been considered, such as the EnCase Academic Program (Guidance, 2014). This is a learning and development programme, where the students study towards the EnCase certification. The core units are driven by the EnCase software, one of the main tools used by the digital forensic investigators in practise. This option could enhance the student experience (Foskett and Maringe, 2010) and potentially the practical expertise of the graduates. However, the planning of this option has not been implemented yet due to cost limitations and commitment issues.

However, we have employed and run an in-house expert witness training (BondSolon, 2012) that prepared the students in writing expert witness reports and presenting their evidential findings in court. This is offered as part of a third year core module in two full days training and the students receive a certificate of attendance.

The feedback that has been received by the students was very positive throughout the years the training was offered to them. They completed evaluation forms, where they used expressions such as: "This was brilliant", "Absolutely awesome", "We should have done this earlier in the course, loved it".

The training required an expert to deliver it with experience in writing effective expert witness reports and presenting evidence in court. An academic may have the expertise in a certain field; however she may lack the practical involvement in specific areas of the discipline required to enhance the student experience.

The expert witness training is a good example of an attempt to adopt professional training as part of a digital forensics undergraduate programme. It involves an external delivery body for the course and offers a professional certification to the students that can potentially enhance their employability in the future. It appeared the course adapted well to the host module and enhanced the student satisfaction.

The students receive a vocational qualification mostly required in the industry to their benefit, additional to their University degree and are professionally prepared for presenting their findings in court; an important component for the digital forensic practitioner.

There is no evidence whether the expert witness training has enhanced the student employability as there is no data to measure this yet.

6. Conclusion

The students who study digital forensics perceive academic qualifications from academics in the University and then apply their knowledge in practise through a practical role industry. As Ramsden (1991) states about the aim of teaching is to make student learning possible.

Therefore, the requirements of the current practises need to be met in a way that advances and enhances the learning experience in the University environment. The implementation and adoption of different techniques might be a necessity.

A number of theories and models are involved into our teaching and learning approach for educating and training digital forensic investigators. Throughout our arguments we refrained from referring to commercial vendors and product training as we deem this to be the by-product of a proper education. Instead, we use an integrated learning and teaching strategy that addresses the specific requirements in the field of digital forensics, taking into consideration the requirements and needs of our industrial partners and contributing to the achievement of the intended learning outcomes of the programme. Training is not education but proper education does involve an element of training. It is in the best interest of the commercial vendors to continue to provide professional accreditation through the academically accredited programmes and it is in the best interest of Universities to continue using experience-based learning.

Kubler and Forbes (2005) describe the different student employability profiles; cognitive skills, competencies, technical abilities are among other qualities the employers may be looking for in an individual. The list also includes though practical elements and vocational courses, immediately raising the need for the adoption of professional certifications and accreditations in higher education in general.

The different frameworks, models and theories discussed in this paper around teaching, learning and assessment should be considered and applied, and perhaps more models and approaches should be examined and reviewed in the future. Our hope is that all scholars can develop an independent learning culture as they are being exposed to a challenging learning environment and a strict learning culture.

Our approach involves a series of lectures and seminars supported by a series of assessed controlled experiments in specialised research laboratories and a series of workshop programmes including internal and external practitioners to assist experiential learning. A research led directed reading programme, participation to research activities and independent study are imperative to promote critical thinking.

Digital forensics is a relatively new discipline. There is not enough guidance or research with established results yet to define this need in the field. However, based on the current findings it appears that certifications could develop an independent learning culture for the students that could improve their employability.

Our ongoing research in this area aims to identify the academic and career benefit of professional qualifications in digital forensics for students in Higher Education. The study will be strengthened by collecting information and analysing trends from recent graduates and current students in order to measure their career progress with and without professional qualifications.

References

Atherton J S, 2011, Learning and Teaching; SOLO taxonomy [On-line: UK] Retrieved <11/05/2012>, from http://www.learningandteaching.info/learning/solo.htm

Biggs J., 2002, Aligning teaching for constructing learning, Retrieved <11/05/2012>, from www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=477

Biggs J. and Collis K, 1982, Evaluating the Quality of Learning: the SOLO taxonomy New York: Academic Press. SOLO taxonomy.

Biggs J. and Collis, K., 1989, 'Towards a model of school-based curriculum development and assessment using the SOLO taxonomy.' Australian Journal of Education, 33(2), 151-163.

Biggs, J. and Tang C., 2007, Teaching for Quality Learning at University, 3rd ed., e-book, Accessed <11/05/2012>, <http://DERBY.eblib.com/patron/FullRecord.aspx?p=332709>.

Bloom Benjamin S. & David R. Krathwohl, 1956, Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook 1: Cognitive domain. New York , Longmans.

BondSolon, 2014, Expert Witness Courses, Internet Source: http://www.bondsolon.com/expert-witness/courses. Last accessed <20/12/2013>

Boud D., 1995, Enhancing learning through self-assessment, London: Kogan Page

Boud D., Keogh R., Walker D., 1985, Reflection: Turning Experience into Learning, USA: Routledge

Brown G., 1997, Assessing Student Learning in Higher Education. London: Routledge.

Brown G. 2001, Assessment: a guide for lecturers York: LTSN Generic Centre, Retrieved <11/05/2012> from http://www.heacademy.ac.uk/resources/detail/assessment/ assessment_series

Caloyannides, M.A., 2001, *Computer Forensics and Privacy*. Artech House Inc

Coffield F., Moseley D., Hall E., Ecclestone K., 2004, Learning styles and pedagogy in post-16 learning, A systematic and critical review, Learning and Skills Research Centre, Retrieved <12/05/2012>, from http://lerenleren.nu/bronnen/Learning%20styles%20by%20Coffield%20e.a..pdf

Covic T. and Jones M.K., 2007, Is the essay resubmission option a formative or a summative assessment and does it matter as long as the grades improve?, Assessment & Evaluation in Higher Education, Volume 33, Issue 1

Cruickshank D.R., Metcalf K., Jenkins D.B., 2009, The Act of Teaching, 5th ed., McGraw-Hill Higher Education ISBN: 0073378399

Elbow P., 1994, Will the virtues of portfolios blind us to their potential dangers? In P. Stillman (Ed.), New directions in portfolio assessment (pp. 40-55). Portsmouth, NH: Boynton/Cook Publishers.

Ennis R. H., 1993, Critical thinking assessment, Theory into Practice, Vol. 32 (3), 179-186.

Forehand, M., 2005, Bloom's taxonomy: Original and revised.. In M. Orey (Ed.), Emerging perspectives on learning, teaching, and technology. Retrieved <11/05/2012>, from http://projects.coe.uga.edu/epltt/

Foskett N. and Maringe F., 2010, Globalization and Internationalization in Higher Education: Theoretical, Strategic and Management Perspectives, Continuum International Publishing Group, ISBN-10: 1441132775

Fry H., Ketteridge S.A., Marshall S., 1999, Handbook for teaching & learning in higher education: enhancing academic practice, London : Kogan Page

Harrow A.J., 1972, A Taxonomy of the Psychomotor Domain, New York: David McKay Co

Higher Learning Commission, 2012, The Criteria for Accreditation and Core Components, Internet Source: https://www.ncahlc.org/Information-for-Institutions/criteria-and-core-components.html. Last accessed: 28/02/2014

Guidance, 2014, EnCase Academic Program, Internet Source: https://www.guidancesoftware.com/products/Pages/encase-forensic/academic-program.aspx. Last accessed: 20/2/2014

Irons A.D. and Kostadopoulou A., 2007, Promoting Professionalism in Computer Forensics: who should accredit?", in proceedings of 2nd European conference on Computer Network Defence, in conjunction with the 1st Workshop in Digital Forensics and Incident Analysis, pp.115, Springer

Kolb D.A., 1984, Experiential Learning: Experience as the Source of Learning and Development, Englewood Cliffs, NJ: Prentice Hall.

Kolb D.A., 2000, Facilitator's guide to learning. Boston: Hay/McBer.

Krathwohl D. R., Bloom, B. S., & Masia, B. B., 1964, Taxonomy of educational objectives; the classification of educational goals. Handbook II: The affective domain. New York: Longman, Green.

Kubler, B. and Forbes, P. (2005) Student Employability Profiles: A Guide for Employers. London: The Council for Industry and Higher Education.

Palmer G.L., 2001, *A Road map for Digital Forensics Research*, in *1st Digital Forensics Research Workshop*.

Ramsden P., 1991, Learning to Teach in Higher Education, Routledge, ISBN-10: 0415064155

Rapp C. Aristotle's Rhetoric. Stanford Encyclopedia of Philosophy, 2010. Last accessed <20/12/2013>, from: http://plato.stanford.edu/entries/aristotle-rhetoric/

Schmidt HG, 1993, Foundations of problem-based learning: some explanatory notes, Medical Education 27 (5): 422–432

Schön D.A., 1991, The reflective practitioner, How professionals think in action, Ashgate Publishing Limited

Schweitzer D., 2003, *Incident Response: Computer Forensics Toolkit*. Indianapolis: Wiley Publishing Inc.

Simpson E.J., 1972, The Classification of Educational Objectives in the Psychomotor Domain Washington DC: Gryphon House.

Tudor I., 1993, Teacher roles in the learner-centred classroom ELT Journal 1993 47: 22-31

Van den Berg I., Admiraal W., & Pilot A., 2006, Designing student peer assessment in higher education: Analysis of written and oral peer feedback. *Teaching in Higher Education*, *11*(2), 135-147.

Wright L., & Jones A., 2004, Good Practice and Innovation in the Assessment of Students in Built Environment Subjects in Higher Education. Retrieved <28/02/2014> *Avalaible at* cebe.cf.ac.uk/learning/sig/assessment/pdf/AssessingStudents1to9.pdf

Knowledge Accessibility and Cyber Macht

Leigh Armistead[1] and Scott Starsman[2]
[1]Peregrine Technical Solutions LLC, Yorktown, USA
[2]Avineon, Inc., Alexandria, USA
larmistead@gbpts.com
sstarsman@avineon.com

Abstract: This paper continues the process of laying the groundwork for a new comprehensive academic theory on Cyber Macht (Cyber Power). The first paper served to lay a foundation for this proposed theoretical construct, while the second paper focused on global influence, namely the fact that the ability to influence events no longer resides primarily at the national or even governmental level. Specifically in that latter effort, the authors stressed that small groups of people and even individuals with a potent message and a well-chosen audience were able to broadcast their message, excite a population, and even initiate an attack. This paper focuses on Knowledge Accessibility, notably that information that was previously difficult to find or access is now fully indexed, searchable, and downloadable over the Internet. Even information that is intended to be restricted can often now be obtained through direct, targeted action. This new theory that the authors are endeavoring to develop, starts with the fact that communication paths and changes in connectivity have allowed power to become globally distributed. The root cause of this shift is the enormous increase in access to information for all people around the world. Extra-national groups and individuals can take advantage of this dramatically increased connectivity and influence events far beyond the previous normal range. It is our desire that this new theory will reference power and influence operations including the ability to shape the information in this new cyber era, while still protecting the assets and maintaining the ability to get accurate and relevant information from multiple sources and channels. Likewise throughout this entire series of papers, the authors' concept has been motive-neutral … i.e. we do not assume that Cyber Macht will be used benevolently; instead it is simply a force for good or for evil, that can be harnessed to influence national policies, actions, and perceptions. Finally, the authors of this paper believe that the only way to have a comprehensive approach to the development of theory on Cyber Macht, is to involve not only the academic community but also the various strategic centers of excellence across the DoD and State Department, which we have endeavored to accomplish in this research.

Keywords: power, theory, strategy, academic, comprehensive

1. Introduction

The world has changed immensely over the last three decades with respect to access of knowledge. Whereas information was previously more centralized, restricted, and controlled, that is no longer the case in the world of Cyber Macht. For example, satellite imagery that was considered classified 50 years ago is now viewed as inferior to Google Earth imagery, which is available to all with a click of the mouse. Likewise, the ability to move data via the internet or even more pedestrian methods such as a memory stick, has transformed the manner in which the world operates. No longer is key information stashed in a safe, where only a limited few have access, but instead a tremendous amount of knowledge is now being posted to the cloud or other readily accessible databases. All of these changes taken together have turned the access to information completely on its head, as will be shown in this paper. In the interest of increased connectivity and productivity, information is instead made readily available to all with the correct URL, password, or a credit card. Unfortunately these relaxations of the old checks and balances that existed previously through physical security measures, have long since evaporated and we now live in world of near instantaneous and constant access to knowledge as desired, but not necessarily as required. A good example is the National Security Agency (NSA) and Edward Snowden, which is detailed in our enclosed Case Study. This vignette demonstrates that unlimited and lax access to critical information may be convenient but also concentrates critical information, dramatically increasing the value of penetration to our adversaries. To that end, it is the authors intent in this paper to detail how the global power structure has changed with this increased and near universal access to knowledge and data and to make recommendations for protecting this information and to move forward.

2. Why is access to information changing?

Information is and always has been a somewhat nebulous term, but in this new era it possesses a capability that is now considered crucial to the success of American national security. However, there are still many questions regarding how to most successfully utilize this element of power. Because the use of data crosses so many boundaries within the interagency processes, it is often very difficult to quantify exactly what constitutes an information campaign. That is because other organizations within the federal bureaucracy, such as the

State Department, which have traditionally concentrated on diplomatic efforts to support American interests abroad, are now instead being asked instead to facilitate strategic information or communications activities around the world. This kind of tasking often belies the normal chains of communication and day-to-day procedures for the State Department, especially with the demise of the USIA activity in 1999. More often than not, the most recent attempts to conduct strategic high-level information activities in the United States at the Executive Level are subjugated to a more tactical set of options that are normally conducted by the Department of Defense (DoD) as part of its standard set of operations.

The contemporary world is now in a period of manipulated images, where nations, groups and individuals all attempt to manage the messages that are received. Information campaigns have been advocated and theorized to be conducted in a very similar manner, whether one is 'selling' merchandise like a soft drink or a threat to national security like weapons of mass destruction. The whole idea which the acolytes of informational power advocate with respect to Information Operations (IO) is to influence the mind of the consumer or the public, to get them to believe in a product or cause. The nature of power has changed and information and perception management figure prominently in any effective exercise or power strategy. To date, the authors have not witnessed this revolution in the use of information, notably this transformation of the structure of power around the world. Broad, strategic governmental adoption of these approaches has not taken root on a global level although there is widespread evidence of tactical use With regard to massive fundamental changes envisioned within the United States federal government, particularly with respect to influence campaigns, the authors do not believe they have occurred and, in many cases, our research indicates foreign policy operations are still conducted using traditional military and diplomatic methods.

2.1 Initial IO theories

These transformational ideas of how one can use information are crucial, because as the events of 11 September 2001 indicate, military, political or economic power are often simply ineffective in dealing with these new kinds of threats to the national security of the United States. The aforementioned terrorist attacks were a blow to the American public and government images that affected the perceptions of many Americans. In addition, as this paper will indicate, the fear produced by these terrorist acts can only be defeated by using a comprehensive plan, in which information is a key element. For, as both Arquilla and Ronfeldt (1999) argue, in the future it will be networks that will be fighting networks. Operation Enduring Freedom and Operation Iraqi Freedom were such examples, where networks in the form of information campaigns fought networks made up of perceptions, and the side that ultimately emerged as the victor, was the one that can best shape and influence the minds of not only their adversaries, but their allies as well (Advisory Group on Public Diplomacy for the Arab and Muslim World, 2003).

The shift from the industrial age to the information environment may mean that the United States will not remain the dominant player in the political arena. For example, Arquilla and Ronfeldt (1997a, 1997b) write that nation-states are losing power to hybrid structures within this interconnected architecture, where access and connectivity, including bandwidth will be the two key pillars of any new organization. They posit that truth and guarded openness are the recommended approaches to be used in both the private and government sectors to conduct business, and in their opinion, time zones will be more important than borders. It will be an age of small groups, using networks to conduct 'swarming' attacks that will force changes in policy. Key features include:

- Wide open communication links where speed is everything
- Little to no censorship, the individual controls his own information flow
- Truth and quality will surface, but not initially
- Weakening nation-states and strengthening networks

2.2 Changes in IO theory in the post 9/11 environment

In the last 12 years, many of the earlier theoretical concepts of IO have been tested in conflicts and cyber active engagements, with a number of ideas being validated, but also quite a few that have turned out to not be correct. It is in this chaotic early stage of a new era, when the disconnect between theory and reality is perhaps greatest, and in particular the inability to match a strategic theory to the changes in the power structure of the federal government are most noticeable and very evident in the United States.

The changes that are mentioned in Noopolitik were truly revolutionary and describe a profound shift in the nature of power. Unfortunately, this transformation has not been translated from a strategic concept to tactical actions (Kuusisto, 2004). Thus the intent of this paper is to try to fill that void because, as related in Rosenau (1998, p. 33) "[a] new lexicon is needed for this purpose…there is a huge gap between our sense of profound transformations and our ability to grasp them from a huge shortage of the tools needed… our vocabulary and conceptual equipment for understanding the emergent world lag well behind the changes themselves."

So while Soft Power and Noopolitik may have struck a chord within the Department of Defense and a number of federal agencies at some point, to date, none of these attempts to develop an overall encompassing IO academic theory for what is happening with regard to information has developed into an accurate and comprehensive construct across the United States as a whole. Even the authors of Noopolitik themselves - Arquilla and Ronfeldt - note as much in a recap to their book The Promise of Noopolitik, published eight years after the original publication of Noopolitik (Arquilla and Ronfeldt, 2007)

Their initial enthusiasm for this theoretical construct has been dampened considerably not only by the events of 9/11, Operation Enduring Freedom and Operation Iraqi Freedom, but also by the way the Internet and the intellectual community have evolved in the last decade. The hopeful optimism of the 1990s with regard to the World Wide Web and the Internet has instead turned in the last few years to the awful realization that given the power of information, many individuals and groups have instead used this new technology to their advantage, whether for their political, financial or social gain (Ibid). Likewise Arquilla and Ronfeldt also admit in their postscript published in 2007 that the early promises of a global community are instead overwhelmed by the day-to-day events, which tend to mitigate the promise of revolutionary change. Although they still believe that Noopolitik is an idea for the future, and while they remain optimistic, they are also dismayed as well by a number of trends as shown below:

- Notions like Noopolitik are gaining credibility, but all too slowly
- Soft Power lies behind them all, but the concept needs further clarification
- Activist Non-Governmental Organizations representing global civil society are major practitioners of Noopolitik, but the most effective may be the global network of jihadist
- American public diplomacy would benefit from a course correction (Ibid)

What is interesting about this concept, is that while it may be the most difficult to achieve from an IO perspective, it may also offer the most promise. One only has to review the four definitive airplane bombing surveys of World War II or Vietnam, to quickly realize that military power often does not translate at all into desired outcomes. Clodfelter (1989), a retired Air Force officer, said as much in his book, on the air campaign in North Vietnam. As these well-researched and documented official reports from the US Air Force alluded to, massive bombing operations did not necessarily and in many cases did not translate at all to shifts in the affected government or populations attitudes. As one veteran (and perhaps jaded) military officer once quipped, "[i]f the only tool you have is a hammer, every problem looks like a nail" (Hubbard, 2004). So too is the case of trying to take military power, in this case aviation assets, and translate into recognizable outcomes.

3. Technology changes that have eased access to information

The ability to access and move data has changed dramatically in the last two decades. Everyone with a computer or smart phone and the Internet can quickly view information that previously one had to physically go to a library, research center, laboratory or other knowledge centers to view and use this material. However, with thousands of databases, portals, websites and other online sharing sites, one can now access this information without leaving one's own home. This is both a wonderful technological feat but also a nightmare in trying to protect the data, which is why the authors advocate a policy to mark and tag the information that is crucial so it can be protected. Information with lower importance is not protected. That may sound harsh, but it is the authors' opinion that, in this world of decreasing resources and increasing cyber threats, it may be one of the few practical alternatives. One cannot protect everything, nor should one ignore the law of diminishing returns and really try. Not all knowledge is classified or proprietary in nature and therefore does not need to be protected.

Other academic sources envision a restructured concept of power in which information plays a more vital and central role. For, if information is now accepted as an element of power, should there not be dramatic changes as well from previous theories? Is the power of information new or different, as some advocates believe, or has information always been an element of power, but it could never be properly utilized? Said in another way, has information always been an element of power and it is only now that technology can manage and harness that power? Critics of this new view of power have argued that because the world access to the Internet is not universal, this new technology cannot truly change global politics. Wriston (1997) notes that while maybe this is true, it is irrelevant. The standard has been set, and the benchmark is high, for these new views of information flow must be understood and respected. In fact, the percentage of overall access and connectivity to the internet are on the verge of exploding as the combination of cellular technology and cheaper interface devices proliferate.

However, the question is whether access to technology necessarily equates to greater power to a group or nation. Once again, the short answer is that it depends. As Treverton (2001) relates,

> *State power can be conceived at three levels: (1) resources or capabilities, or power-in-being; (2) how that power is converted through national processes; (3) and power in outcomes, or which state prevails in particular circumstances. The starting point for thinking about—and developing metrics for—national power is to view states as "capability containers." Yet those capabilities—demographic, economic, technological, and the like—only become manifest through a process of conversion. States need to convert material resources into more usable instruments, such as combat proficiency. In the end, however, what policymakers care most about is not power as capability or power-in-being as converted through national ethos, politics, and social cohesion. They care about power in outcomes. That third level is by far the most elusive, for it is contingent and relative. It depends on power for what, and against whom? (ibid, 2000, p.11).*

4. Case study

The history of information relative to the attainment national objectives has a long history and most of it involves national powers using information to support their objectives. Even during the Second World War, the media was largely cooperative with the government and media in protecting critical national (and sometimes personal) secrets. However, during the Vietnam War, activist American journalists grew increasingly dissatisfied with their role as the unofficial mouthpiece of the Government and began extensive and broad reporting of information at odds with governmentally-established national objectives. This negative reporting was a key factor in the erosion of public support for the US involvement in Vietnam that ultimately led to the withdrawal of US forces and the fall of the South Vietnamese government (Oram 1993, p.9). Over the next twenty years, the media and the US government established a new equilibrium with the media occupying a self-appointed "watchdog" role over the government.

As the US entered the Information Age, big-media began being challenged for supremacy in its role as news and entertainment providers. Individual reporters and bloggers began reporting and commenting on current events. Communities of interest began to form around these individual reporters as they pursued storylines of particular interest to their followers. One such reporter was Glenn Greenwald who began blogging in 2005 and focused on issues involved with privacy and national surveillance. His attention to privacy and arguments against national surveillance attracted and, perhaps, inspired a young IT technician, Edward Snowden, to entrust him with highly-classified documents that ultimately compromised a multi-billion dollar global surveillance program. While there has been and continues to be much discourse on the ethics of his actions, which is not the purpose of the discussion here. Rather, the authors will examine this case to understand and highlight the enormous challenge in securing information and preventing its unintended disclosure.

As the world entered the digital age, paper documents, products, and processes were digitized and made available via computer networks. This process began largely with text documents and expanded to include pictures, presentations, video, audio, geospatial representations, etc. In the recent past, the signal medium was different depending on the information type. For example, voice communications were transmitted over analog telephone lines, video was transmitted over serial digital lines, and text traffic used TCP/IP. Infrastructure and communication efficiency has driven almost all modern information transmissions to TCP/IP networks. This approach reduces costs and provides opportunities for information integration that were not formerly possible.

However, like the rest of the authors' cyber theory, the power of information convergence is motive-neutral. Just as the end-users benefit from the concentration of information so can those with motives counter to the content-owner's interest? While large organizations and governments have gained great understanding and power by storing and indexing all the data collected, it also creates tremendous potential for adversaries to gain access to and exfiltrate information that was formerly too difficult to obtain, store, and search.

Such was the case with Edward Snowden. He was able to search for, locate, download, and remove up to 1.7 million documents [Zakaria and Strobe 2013]. His revelations did immense damage to the technical capabilities of the agency and have caused major damage to the reputation and trust of the United States and companies doing business there. US citizens are suspicious of their own government's capabilities and motives and public trust in the US government has reached an all-time low, lower than even during the Watergate era [Pew 2013]. The inability of an agency such as the NSA to fully appreciate the vulnerability of this expensive and valuable capability to such an insider threat highlights the limits of understanding of the power and threat that even a single well-positioned individual can pose in this new age. Despite extensive logging and monitoring, Mr. Snowden was able to collect a large library of highly-classified documents and remove them from US Government control.

To view this incident as an isolated incident perpetrated by a single traitor or idealist is to miss the central point of Cyber Macht: in this world even individuals have the power and capacity to act on a global level and exert influence at a level on par with nation-states. This creates a new dynamic whereby nations must erect defenses against other nations (which is at least relatively well-understood and for which there exists basic diplomatic rules that govern behavior and consequences) but also against small groups and individuals with widely varying motivations and goals. These threats may well exist amongst the most trusted participants and are capable of tremendous damage. Erecting defenses against a threat of this nature is a virtually impossible task and requires a completely different approach to mitigation and defense.

5. A proposed way ahead

As has been alluded to throughout this paper, the authors believe that there is a middle road ahead, between a wide open unfettered access to all information, and a locked down retreat to the past. The authors do not believe the latter is an option, with today's society and its desire for 24x7 connectivity, so what we are proposing a solution that utilizes processes and a combination of hardware and software to protect the key information that is required. The solution that we advocate is a flexible option, one that allows organizations to adapt to changing circumstances as they arise. It is by no means to be all inclusive but instead this proposal is meant to serve as an academic model, a theory to be tested and examined in the real world. The key concept of our thesis is that not all information is important, and that there is a time element, perishability to the importance of any data. So in order to decide what and how one is to protect the knowledge and ensure access for those as desired or required, then one must first analyze the data and mark or tag that information, and then it can be stored or transferred in the proper manner. This metadata has been known for a long time and it is not a new concept, but the authors believe that, by adding the time element, a test set can be built and then be evaluated.

In this manner, the user can decide which information is important … and for how long. The latter is key, because one cannot store all data forever, nor can one retrieve legacy information, once its storage mechanisms are no long in common use (i.e. floppy disks, zip drives, etc.). After this first step, one needs to determine access, and this is where a combination of hardware, software and processes will be key. If one is using a wedding cake model for depiction, then the most important data is at the highest level, and takes the most effort to access. The problem that arises is that once a trusted entity has been granted access, then relinquishing those permissions or removing their ability to log in to these networks or databases is more of an administrative process which is often not given the importance deserved. For example in the DoD, the certification revocation list for PKI which is often woefully behind and out of date. While many organizations will ensure that they get keys or entry cards from departing employees, deleting access to networks is not always done. Even worse as in the cases of Edward Snowden or WikiLeaks, these acts of downloading huge amounts of data were done by current employees, in apparently good standing with their respective organizations. So unless one was monitoring traffic or reviewing the log files, or watching for uncharacteristic behavior, it is very difficult to stop this type of insider behavior. There are some hardware and software tools and applications that advertise that they can solve this issue, but in the authors' opinion it is more a

management problem that can be better solved by a rigorous information protection methodology that treats the important data as identified earlier with the respect deserved. For example, all Top Secret used to be marked with red folders and kept in safes where only certain people had access. That same model can be applied in the cyber world, where data is segmented and protected by its metadata tags, i.e. the more classified and timely, and the more layers of defense. In addition, management needs to review individual's pattern of data usage more closely, to be aware of who is accessing information that is outside of their work tasks or areas. One will not employees to physically enter restricted areas, nor should employees be allowed to randomly troll networks as desired. More restrictions on access should be added to organizations' information policies, to include 'need to know'. In addition, when employees are accessing data this should be reviewed and the movement of files across the network should be monitored and include the time of day or day of the week when distinct differences in behavior are noticed. The key to these increased management involvement in the protection of the knowledge and limiting access is to develop a methodology that works for your particular organization. But if protecting the data is one side of the coin, the power gained by the increased ability to use this knowledge to increase one's ability to influence events on a much broader scale. This can be both a good and bad thing, depending on your viewpoint.

6. Summary

The information genie is out of the bottle and will never be put back in, so one must know concentrate on how to protect those assets that are considered most valuable. The authors have presented a series of ideas that combine hardware, software and processes, so that an organization can protect that data that has been deemed critical (metadata). We also discussed in our short case study, a recent example of knowledge accessibility gone wrong, in which a single individual gained a tremendous amount of power (Cyber Macht), with the release of so many formerly classified documents. In this act, Edward Snowden radically transformed the relationship of the NSA with the American public, and also the United States with its allies. He conducted Cyber Macht in the truest sense, giving access to the whole world of previously classified knowledge, upending a number of paradigms within the intelligence community. While the authors do not agree with his actions or decisions in the slightest, he is however an easy example that one can use to site the dangers inherent in making this mind of information readily available to all. As former Naval officers who each served a career in protecting the United States from enemies both foreign and domestic, we hope that the lessons learned as a result of this unfortunate incident can be the impetuous for a stronger and more robust set of policies and procedures to protect key information in the future.

References

Advisory Group on Public Diplomacy for the Arab and Muslim World. Changing Minds, Winning Peace: A New Strategic Direction for U.S. Public Diplomacy in the Arab and Muslim World. 1 October 2003.

Arquilla, J. and D. Ronfeldt. "A New Epoch - And Spectrum - Of Conflict," in Athena's Camp: Preparing for Conflict in the Information Age, ed. John Arquilla and David Ronfeldt (Santa Monica, Ca: RAND, 1997), 1-22.

_____. "Looking Ahead: Preparing for Information-Age Conflict," in Athena's Camp: Preparing for Conflict in the Information Age, ed. John Arquilla and David Ronfeldt (Santa Monica, Ca: RAND, 1997), 439-501.

_____. The Emergence of Noopolitik: Toward an American Information Strategy. Santa Monica, CA: RAND, 1999.

_____. The Promise of Noopolitik. First Monday. 20 July 2007. Accessed at http://www.firstmonday.org/issues/issue12_8/ronfeldt/index.html on 23 January 2008.

Kuusisto, T., R. Kuusisto and E.L. Armistead. "System Approach to Information Operations", 3rd European Conference on Information Warfare and Security, Royal Holloway, University of London, 2004, pp. 231 – 239.

Oram, S. S., "Telling the Story: Impact of Military-Media Relations on the Operational Commander", Naval War College, Newport, RI, Feb 22, 1993.

Pew Research Center, Public Trust in Government: 1958-2013, http://www.people-press.org/2013/10/18/trust-in-government-interactive/, October 18, 2013.

Rosenau, J.S. "Global Affairs in an Epochal Transformation," in The Information Revolution and International Security, ed by Ryan Henry and C. Edward Peartree, CSIS, Washington, DC, 1998, p. 33.

Tabassum Zakaria and Warren Strobe, "After 'cataclysmic' Snowden affair, NSA faces winds of change", Reuters, 13 Dec 2013.

Treverton, G.F. and L. Mizell, The Future of the Information Revolution in Latin America: Proceedings of an International Conference. Santa Monica, CA: RAND (CF-166-1-NIC), 2001.

Wriston, W.R. "Bits, Bytes and Diplomacy." Foreign Affairs 76 (September/October 1997): 172-182.

A Strategic Approach to Managing Security in SCADA Systems

Mehdi Asgarkhani[1] and Elena Sitnikova[2]
[1]Department of Computing, CPIT, Christchurch, New Zealand
[2]School of Information Technology and Mathematical Science, University of South Australia, Adelaide, Australia
Mehdi.Asgarkhani@cpit.ac.nz
Elena.Sitnikova@unisa.edu.au

Abstract: The internet was started as an open distributed system for researchers and academics without much concern for cyber security related issues. However, it gradually evolved into a complex network as a platform for connectivity of businesses and communities across the world. Today, the Internet not only brings together businesses and communities but also hosts complex electronic and engineering systems (such as Supervisory Control and Data Acquisition – SCADA – systems) deployed to manage critical infrastructures. The Internet has become the transporting method for high volume of critical information worldwide. Securing sensitive information and safe-guarding operations of critical infrastructure management systems is a worldwide preoccupation. SCADA systems are put in place worldwide to manage not only critical infrastructures but also complex operations that play a vital role in managing scarce natural resources and the environment. This paper examines research on both SCADA systems and Internet security and provides an overview of key issues and recommends a strategy framework for protection of SCADA systems.

Keywords: SCADA systems, cyber security, strategic security management, information security

1. Introduction

The internet which was originally designed to be an open and distributed environment where security and mutual mistrust was not a primary concern, has evolved into a global network with access available to almost anyone. What's more, critical systems such as Supervisory Control and Data Acquisition (SCADA) systems historically operating stand alone are now connected to the Internet for more efficient communication of information. Cyber security has become a critical issue. Table 1 demonstrates the evolution of the Internet throughout its life.

Information is a critical asset for many companies and is part of the core business for many new companies, (Mellado 2012). In case of SCADA systems critical information generated by sensory components is used to make critical decisions. Vulnerabilities of the Internet and weaknesses of SCADA systems (which can lead to compromises of information security) have existed for a long time. The responsibility for many of these vulnerabilities must be directed towards the companies who are too keen to release fundamentally unsecure software with a 'patch later' attitude (Hayes 2012). The profile of those who initiate attacks (cyber-criminals) can vary significantly, ranging from a shy introvert, smart professional, organised gangs with significant funding to employ smart hackers to government funded groups who have access to state-of-the-art equipment. Their motivations are equally diverse, financial gain, revenge, political, ego and curiosity. At the same, the risk of attacks from the inside of the organization cannot be overlooked. Overall, consequences of security compromises related to SCADA systems can be serious – as they often control and manage critical regional and national infrastructures.

Motivated by factors outlined above, this paper investigates broader cyber security and information security issues (related to the broader internet) with a focus on strategic systems such as Process Control Systems (PCS), Industrial Control Systems (ICS), Distributed Control Systems (DCS) and SCADA. The methodology for this investigation concentrates on review of previous research on relevant topics and analysis of case studies that relate to both broader security considerations and specific SCADA security management.

The paper introduces SCADA systems followed by analysis of vulnerabilities and sources of attacks. After putting issues related to SCADA security into broad categories, it introduces a strategic approach for addressing and managing cyber security issues in SCADA systems.

Table 1: The evolution in use of the Internet

Phase 1 (1960s-1970s)	A network with limited use by academics and researchers.
Phase 2 (1980s – TCP/IP)	The use of the Internet is broaden to most universities and professional institutions worldwide
Phase 3 (1990s)	Government organizations and large organizations start using the Internet (and Intranet) for improved efficiency in communication
Phase 4 (Late 1990s)	With personal computers (PCs) becoming affordable, PCs and Internet connections move into houses.
Phase 5 (Late 1990s – Early 2000s)	The use of Internet becomes widespread in businesses (ecommerce and e-business solutions) and alongside that more sophisticated networking and PC solutions are introduced in homes
Phase 6 (Mid-Late 2000s)	The Internet becomes part of day-to-day life with advanced solutions in businesses and connectivity to businesses provided to customers
Phase 7 (Mid-Late 2000s)	Integrating sophisticated industrial control systems such as SCADA into web and Internet solutions for strategic management of critical resources

2. Introducing SCADA

SCADA systems are generally used to automate complex processes where human interference to collect data in assessing overall performance of a system is not possible. Cases of SCADA in managing electricity generation and distribution or in managing water resources are examples of SCADA as strategic technology system with benefits that reach beyond an organization or one community. Traditionally, SCADA was seen as being a technical engineering tool to automate complex processes. However, soon it was realized that what SCADA systems achieve benefit not only engineers but also decision makers within organizations. It moved from being an engineering solution to be become as a key corporate system. SCADA operations are complex as they involve diverse group of people. Becoming aware of the role of each group and the ways in which groups complement each other's skills is a critical aspect of successful operation of SCADA. The diversity of groups of people and cultures involved can also lead to risks in operational issues including cyber and information security. Today, SCADA is not only connected to internal network within organisation, but also to the Internet. Therefore, SCADA is exposed to cyber security risks discussed earlier - similar to any other system connected to the Internet.

Overall SCADA provides authorities with control capabilities such as:

- Accessing quantitative measurements of important processes
- Detecting and correcting errors in a timely fashion – as soon as they are detected
- Measuring trends in the ways in which a system functions over time
- Detecting and addressing critical issues and bottlenecks and other inefficiencies
- Controlling complex and sophisticated processes without need for specialist human resources

Those capabilities mentioned above are provided by combination of functions which are in turn supported by various components of SCADA. They include four key functions: *data acquisition*; *networked data communication*; *data presentation and control*. These functions are performed by various SCADA components such as *sensors and control relays*; *remote telemetry units* (RTUs); SCADA *master unit* and the *communication network*. Another key component of SCADA is the Human Machine Interface (HMI) or (MMI) –Man Machine Interface. As outlined above, SCADA systems have numerous components performing sensitive tasks communicating critical information via corporate networks and the Internet. Consequently, there are significant potential security risks to these systems.

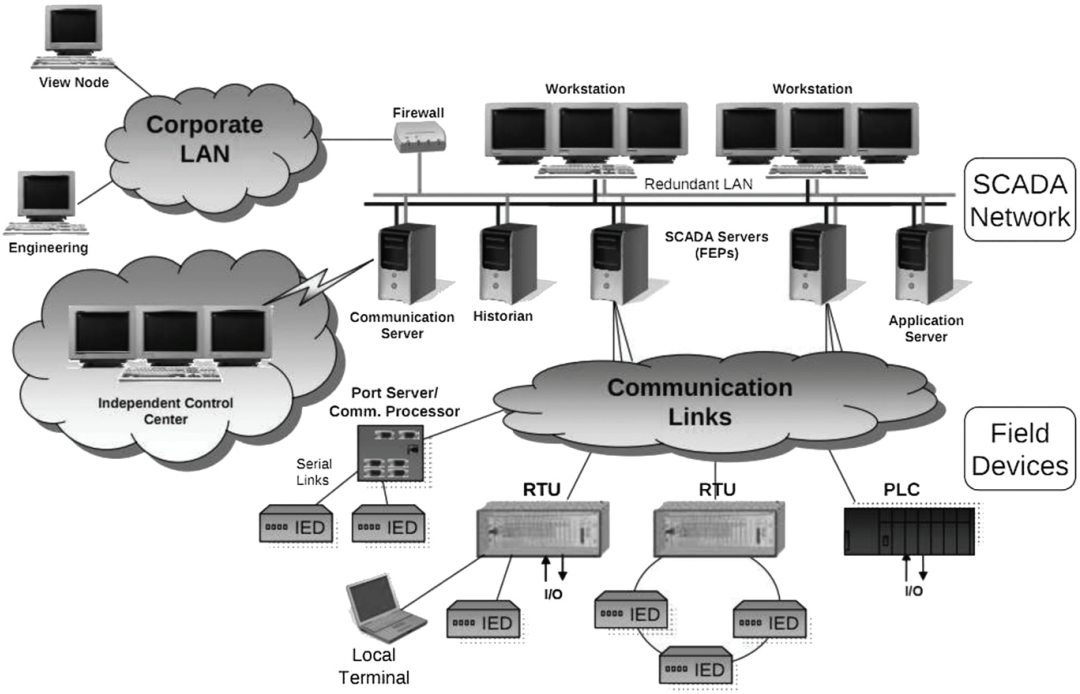

Figure 1: Typical SCADA setup (Source: Pacific Northwest National Laboratory, 2006)

3. Analysis of vulnerabilities

This section examines weaknesses and vulnerabilities related to both the cyber space (the Internet) and SCADA. Vulnerabilities include possible threats and sources (initiators) of attacks. Earlier SCADA systems limited communication of information within the network that managed SCADA within a particular organization. Today, information gathered by SCADA systems can be communicated across wide area networks and the Internet for remote and more effective management of operations. In other words, today, due to openness of networks and connectivity with the Internet managing security of SCADA systems poses new challenges. As outlined earlier, the applications of SCADA range from monitoring lifts and escalators in the London tube system, monitoring wind farms to controlling and ensuring the safety of vital power generation services (BBC, 2004) and many applications in between. The implications of a deliberate attack on any of the mentioned SCADA systems would be severe; consequently SCADA is a prime target for cyber-terrorism.

Vulnerabilities that are specific to the architecture of the Internet are:

- **Connectivity and resource sharing**: the Internet is an open network designed for information sharing accessible globally via a routable IP address. What's more, it is based on a packet-switching framework. These two aspects of Internet architecture allow flooding attacks to occur.
- **Authentication, integrity and traceability**: As there is no inherent authentication in place, IP spoofing where false information is injected into IP packets can be possible. What's more, most routers simply receive and forward packets with no integrity checks. This allows attackers the opportunity to hide their identity and be untraceable.
- **Internet Security**: There are numerous insecure unpatched systems connected to the Internet. Consequently, attackers are able to exploit these vulnerable systems. "We can secure our systems but we cannot force others to do so." - Sachdeva et al (2008).
- **Intelligence and resources asymmetry:** it appears that almost all the 'Intelligence' of the Internet resides at the end hosts. These end hosts are connected via high bandwidth links and routers, it is easy for attackers to quickly overwhelm the resources of the victims system.
- **Lack of centralised control on Internet:** due to the global nature of the Internet, there is no overall central authority responsible for network security. This makes it almost impossible to administer a common global defence strategy.

- **Broadcast concept for data communication:** the broadcast concept implies that other computers should ignore data not destined for them, but they can be modified to listen and even change information in data streams.

According to Adeyinka (2008), some of the more common internet attack methods used by cybercriminals can include:

- Viruses, Trojans and Worms
- System and Boot Record Infectors
- Eavesdropping
- Hacking
- IP Spoofing Attacks
- Denial of Service
- Email Bombing and Spamming
- Phishing
- Man-in-the-middle attacks

Cost of cybercrime to businesses is significant. The Norton Cybercrime Report (2012) outlines that global rate of direct cybercrime losses rose to $110 billion in 2012. Ponemon Institute (2012) Cybercrime Report presents in details the results of a study of views of business leaders and IT professionals located in five countries (United States, United Kingdom, Germany, Hong Kong and Brazil) regarding the effects and cost of cybercrime on business – key findings include:

- Denials of Service (DoS) attacks were considered to pose the greatest risk to organisations.
- U.S, U.K. and Hong Kong listed DoS attacks as most worrying.
- Brazil listed viruses, worms and Trojans as most worrying.
- Germany listed social engineering attacks as most worrying.
- Germany and U.S. reported highest levels of attacks at 82 and 79 respectively.
- Averages of 17% of computers and mobile devices within organisations have been infected in some way.
- Effects on reputation and brand name were of least concern.
- The financial cost to recover after one particular cybercrime attack varied from nearly $300,000 (USD) in Germany to just over $100,000 (USD) in Brazil.
- Hong Kong and Brazil reported the highest number of mobile device (smart phones and tablet PC's) infections at 25% and 23% respectively. Germany reported the lowest number of infections at 9%

The Information Security Breach Impact report (ISBS 2012) indicated that 82% of organisations had experienced staff related incidents, including:

- Misuse of web access
- Misuse of email
- Unauthorized access to systems or data
- Breach of data protection laws or regulations
- Misuse of confidential information
- Loss or leakage of confidential information

With regards to SCADA systems, legacy systems that were once engineered solutions are no longer engineered – that is to say, ideally, security technology should not be bolted on but rather embedded into the application delivery. For the time being, bolted-on security is likely to be seen as the only practical way to tackle the critical security management concerns for SCADA systems. On the other hand, given increased desire for connectivity within various sectors of the industry, SCADA systems that were previously isolated from public networks and the Internet have been integrated into these networks. Consequently, they are exposed to the wider cyber community, hence vulnerable and prone to attacks. The interconnection of old (legacy) SCADA control systems

to the internet, networks, or telephone lines, computer networks and wireless systems has created opportunities for potential terrorists, hobbyist hackers, disgruntled employees and other groups to get access to critical infrastructures (e.g. power grids) and cause enormous damage to communities and economies costing countries billions in losses and damages. Threads to SCADA systems can also originate from unauthorized changes to programmed instructions in the PLC's, RTU's and DCS controllers. This may result to malfunctioning of an infrastructure (Robles, Choi, Cho, Kim, Park, & Lee, n.d, p.18) with significant consequences.

4. Most common sources of threats in cyberspace

Traditionally, systems and technologies (including software and applications) have been seen as the core source of security breaches and compromises. However, recent studies in technology management (Asgarkhani, 2013) clarify that procedural and human factors can be equally important sources of breaches. Nicholson (Nicholson et al, 2013) and Iversen (Iversen, 2004) suggest that between 1982 through 2000 SCADA attacks were 70 percent internal – mostly related to disgruntled employees and mistakes). The rest (30%) represented external sources – such as hackers and cyber terrorists). On the other hand, Eric Byres (Iversen, 2004) presented a different pattern that between 2001 through 2003 70 percent of compromises were external and 30 percent were internal. Nicholson (Nicholson et al, 2013) explains that "it is unlikely that the number of internal attacks had lowered in any way; simply that the number of external attacks had risen so much as to cause these figures." Regardless, the statistics presented above demonstrate that internal sources of compromises are to be treated seriously as they represent a significant proportion of breaches.

Some of the most common sources of attack initiators are:

- *Bot network operators* – They are hackers who instead of breaking into systems take over several systems to coordinate attacks and distribute phishing schemes, spams and malware attacks.
- *Terrorists* - Amongst the most concerning sources of attacks is terrorism. Nicholson and Iversen (Nicholson et al, 2013 and Iversen, 2004) suggest that there are no distinctive SCADA cyber terrorism incidents but they argue that the events of 9/11 suggest that it is possible. In fact in 2001 suspicious patterns of intrusions were detected on the website of a Californian city in Silicon Valley; Mountain View. Further investigation by the FBI highlighted similar cases on other U.S. cities' websites and realised that the intrusions originated from the Middle East and South Asia. Iversen (Iversen, 2004) quotes Dave Sanders (director of the Control System Security Centre at the U.S. Department of Homeland Security) stating that "there's no difference in the way that terrorists are going to get into your networks, the real difference is going to be the intent and the payload, and may be even the consequences. The intent will be specifically directed at targets within your critical infrastructure your control system, or possibly even a substation"
- *State Sponsored Hackers* - Currently, the extent of attacks conducted by these kinds of attackers seems to have been minimal. However, state hackers should be considered to be one of the most dangerous threats to SCADA systems – simply due to easy access to resources including financial means.
- *Criminal groups* (organized crime hackers) – In April 2009 hackers held a pharmacy at ransom claiming to return patient records for the fee of $10 million dollars (Goodin, 2009a and Nicholson et al, 2013)). Organised crime is motivated by money and hence holding companies to ransom by attacking their SCADA systems provides an effective method of achieving outcomes cyber criminals desire.
- *Insider attacks* (Disgruntled employees) - Disgruntled employees are reported (Blau, 2004) to be the most common internal source of security breaches. Since inside attackers may have access to the computer network, they are able to bypass the typical security measures that are in place to protect from outside attacks. What's more, inside attackers in many cases already have authorised access to the network – for instance they can easily connect infected devices such as an external drive into network computers to compromise them.
- *Hobbyists* - Hobbyists do not necessarily possess skills and experience at the level expected from professional cyber criminals. However, they are driven by curiosity or seeking popularity and can cause significant damage to infrastructure if insufficient safeguards are in place.
- *Hacktivists* - Activist hackers or "Hacktivists" can also cause security risks to SCADA systems. Attacks based on political reasoning and motivation has been cited in a number of government website defacements (Goodin, 2008b). Consequently, one can assume that they can be equally capable of attacks on SCADA systems.

- *Unskilled key personnel* - IT personnel who are authorized to access critical components of networks and SCADA systems but are do not possess necessary technical skills and are unaware of right technical processes while operating systems can unknowingly expose systems to cyber security breaches.

SCADA systems may not necessarily be under attack from all sources mentioned above at the same time. What's more, some of the sources mentioned above are more likely to be the cause of security breaches in SCADA systems than others – depending on the type of industry involved. Managing SCADA systems security needs to be based on strategies that take into consideration all factors that are mentioned above.

5. A strategic approach to managing cyber security

The total strategic process for managing is anything but linear (Asgarkhani, 2011). Integrating all the components of the strategic process is cyclic. The key elements of this cycle (as outlined in Figure 2) are:

- **Strategic Analysis for security management –** involves establishing an understanding of the current situation, including: aspects of the environment; current technology infrastructure; potential vulnerabilities; expected outcomes and analysis of implications of breaches.
- **Strategic Choice for security management** – involves the formulation of the strategy itself through understanding various options, evaluating options and making a decision on a suitable security strategy.
- **The Security Management Strategy Implementation** – involves tactical issues such as resource assessment and planning, identifying skills & systems, determining organizational structure including role of people.

Nicholson, Webber, Dyer, Patel, & Janicke (2012, p.10) used cases and facts to highlight that any company or government that uses a SCADA system is exposed to significant risk - should that system fail, or be compromised. The risks include brand damage, loss of revenue, share price reduction and in severe cases and most importantly loss of life. A survey of nearly 700 participants conducted by SANS Institute (http://www.sans.org) suggests that 70% of participants rate the risks to their systems to be from high to severe. The same survey indicates that 33% of participants suspect they may have had incidents. This suggests that the operators of SCADA system operators are aware of the risk to their systems. Robles, Choi, Cho, Kim, Park, & Lee (n.d, p.18) classify threats to critical infrastructure into three categories as natural threats, human-caused, and accidental or technical. However, in this paper a broader (strategic) classification is taken into consideration in order to develop a cyber-security and risk management framework.

Taking into consideration a strategic analysis approach (Figure 2) and reviewing what has been presented in this paper, there seems key issues (factors) in managing cyber security:

- **Technology** – Systems hardware and software
- **People** – who work with and/or manage SCADA systems including those who set processes
- **Processes –** which govern use of systems by people

Some of the most common strategic options and/or solutions to address the technology aspect of security can include:

- **Penetration testing** - These tools and solutions are used to help security professionals to perform tests on systems including SCADA systems to assess if systems are secure against risks and exploits that are known to the community of security management professionals.
- **Honeynets** - Another SCADA specific tool is SCADA Honeynets (e.g. SCADA Honeynet Project, 2003). These are simulation models of SCADA systems that can be put in place to distract hackers from targeting the real system. At the same time, SCADA Honeynets, as simulation models can be used to gather and record statics related to attacks – such as sources of attacks, frequency of attacks and implications of attacks.
- **Simulated Attacks** (War games) - Another technique for managing system security is simulated attacks (Greene, 2008).
- **Access Control** – Discretionary Access Control (DAC) and Mandatory Access Control (MAC) are two different approaches. Commercial systems use DAC in which the resource owner determines who can gain access. The MAC system decides itself who can gain access to the particular resource, much like a security guard.
- **Encryption Systems** – this involves encrypting data with a unique key.

- **Firewalls** – designed to block unwanted traffic either in to or out of a private network.
- **Intrusion Detection Systems** – an IDS system can be hardware or software based and is designed to monitor the network to look for signs of an attack. Some systems monitor and alert while others actually block the attack.
- **Anti-Malware Software and Scanners –** these are software based and designed to seek out and destroy malware (viruses, worms, Trojan horses, etc.) from infected systems.
- **Internet Protocol Security (IPsec)** – provides fundamental security at the IP layer of the TCP/IP protocol. It is a point-to-point protocol in which one side encrypts and the other decrypts using one or more shared keys (Adeyinka 2008).
- **Secure Socket Layer (SSL) –** this is a suite of protocols designed to use key exchange, authentication and encryption to create a secure channel for transmission of data.
- **Honey Pots** – these are decoy servers or systems designed to gather information regarding an attacker or intruder attempting to gain access into your system.
- **Sandboxing** - enables captured potentially malicious code to be executed in a safe and controlled environment.
- **Black Listing -** is a network administration practice used to prevent the execution of undesirable programs.
- **White Listing –** is similar to Black-listing, but this time a simple list of authorised programs is maintained.

Systems and technology related security measures alone cannot guarantee secured operation of systems. Systems are run by people, regardless of robustness of design and reliability of protecting solutions, people related issues can compromise systems. Some of the people related issues that can potentially create an environment where security is compromised include:

- Lack of understanding and awareness of implications of security compromises
- A relaxed culture where system reliability is not taken seriously
- Lack of training for admin staff so they can understand functions and risk implications
- Lack of management training to be aware of value of security and cost of being exposed to risks to their businesses
- Shortage of suitably trained and skilled technical staff who manage the operations of the system
- An environment where teamwork is not encouraged – for instance, different groups such as engineers, IT workers and admin workers due to different work culture and personalities run the risk of not working together towards a shared objective.
- Cultural differences in multicultural environments where culture crashes may also result in teams not working together towards shared outcomes.

People who work with networks and/or SCADA systems can be put into three main categories: Technical staff (Engineers and IT personnel); Admin staff and Management. Suggested course of action for people generally involves training to improve awareness of security risks and processes. For technical staff, possessing the right skills and possible certification is crucial.

The third element of strategic security management is governance and processes. IT governance is essential to ensure that the use of IT (including SCADA systems) is aligned to the strategic objectives of the organisation. Risk management is an integral part of IT governance. Effective governance of operational aspects can often be achieved via putting in place standardized processes to be followed by all groups of people working on SCADA systems (Morris et al, 2013; Slay and Sitnikova, 2009). Standardised processes can often be more effective if they are based on industry standards. Some of the industry standards that are applicable to both managing network security and SCADA systems are (Susanto, 2011):

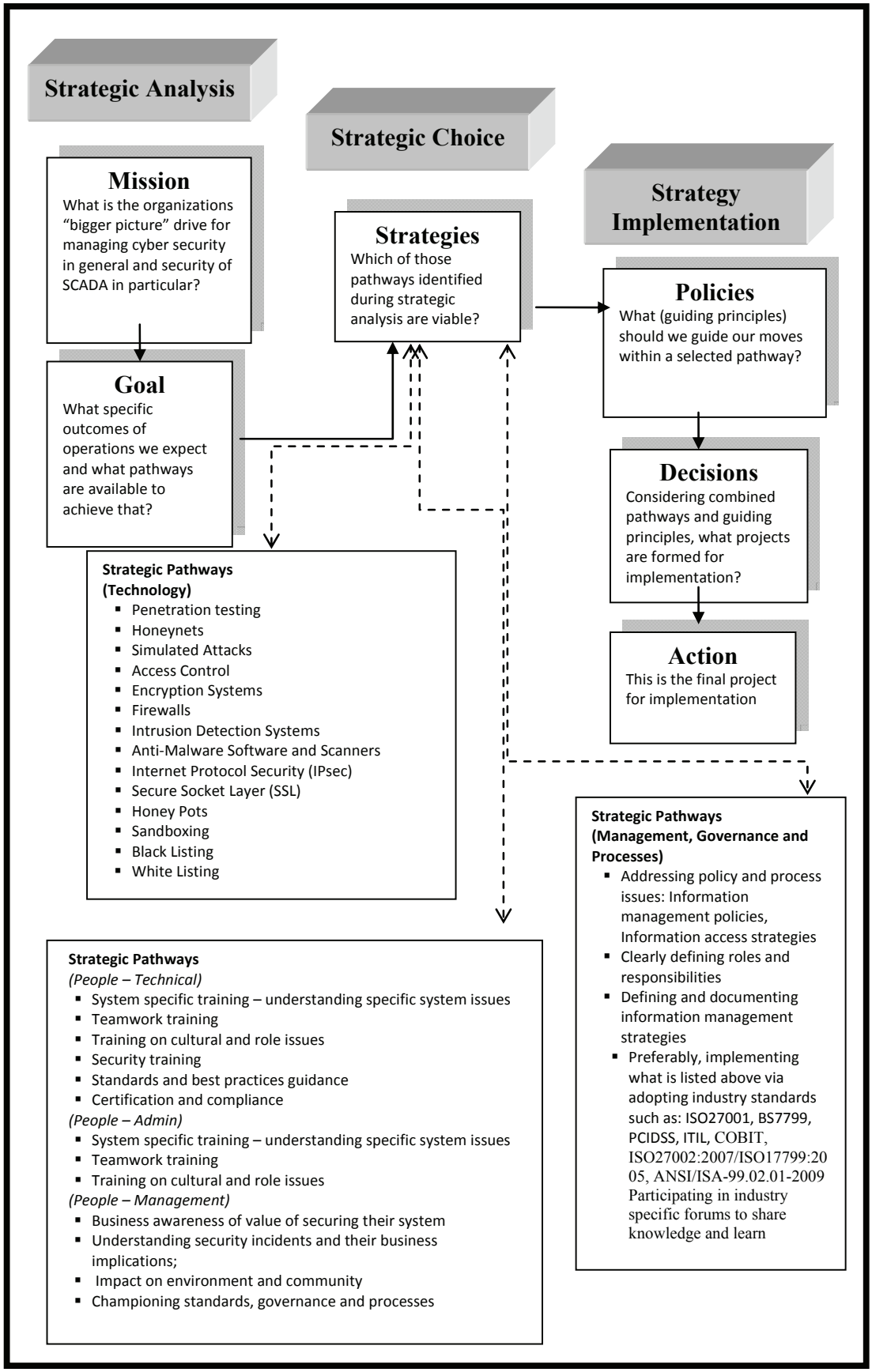
Strategic Analysis
Strategic Choice
Strategy Implementation
Mission
What is the organizations "bigger picture" drive for managing cyber security in general and security of SCADA in particular?
Strategies
Which of those pathways identified during strategic analysis are viable?
Policies
What (guiding principles) should we guide our moves within a selected pathway?
Goal
What specific outcomes of operations we expect and what pathways are available to achieve that?
Decisions
Considering combined pathways and guiding principles, what projects are formed for implementation?
Action
This is the final project for implementation
Strategic Pathways (Technology)
Penetration testing
Honeynets
Simulated Attacks
Access Control
Encryption Systems
Firewalls
Intrusion Detection Systems
Anti-Malware Software and Scanners
Internet Protocol Security (IPsec)
Secure Socket Layer (SSL)
Honey Pots
Sandboxing
Black Listing
White Listing
Strategic Pathways (Management, Governance and Processes)
Addressing policy and process issues: Information management policies, Information access strategies
Clearly defining roles and responsibilities
Defining and documenting information management strategies
Preferably, implementing what is listed above via adopting industry standards such as: ISO27001, BS7799, PCIDSS, ITIL, COBIT, ISO27002:2007/ISO17799:2005, ANSI/ISA-99.02.01-2009 Participating in industry specific forums to share knowledge and learn
Strategic Pathways
(People – Technical)
System specific training – understanding specific system issues
Teamwork training
Training on cultural and role issues
Security training
Standards and best practices guidance
Certification and compliance
(People – Admin)
System specific training – understanding specific system issues
Teamwork training
Training on cultural and role issues
(People – Management)
Business awareness of value of securing their system
Understanding security incidents and their business implications;
Impact on environment and community
Championing standards, governance and processes

Figure 2: The strategy process for security management

- *ISO27001* – This standard looks at the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining and improving documented Information Security Management Systems (ISMS) within an organisation.
- *BS7799* - This standard concentrates on the implementation of ISMS and also incorporates the Plan-Do-Check-Act (PDCA) system and aligns with the ISO 9000 quality standard.
- *PCIDSS* - The Payment Card Industry Data Security Standard (PCIDSS) is an internally adopted standard defined by the Payment Card Industry Security Standards Council. It was created to help organisation process credit card transactions and prevent credit card fraud by implementing strict data controls.
- *ITIL* - The Information Technology Infrastructure Library (ITIL) is a set of concepts and practices relating to Information Technology Services Management, IT development and IT operations.
- *COBIT* - Control Objectives for Information and Technology was created by the Information Systems Audit and Control Association (ISACA) and the IT Governance Institute [ISAKA]. It is described as a set of practices (framework) for IT management. COBIT has 5 main areas; strategic alignment, value delivery, resource management, risk management and performance management.
- *ISO27002:2007/ISO17799:2005* – ISO27001 was discussed earlier as an information security management standard. The ISO27002 series builds on the old ISO17799 standard which in turn built on the old British standard BS7799 (discussed earlier (IT Governance, 2011).
- *ANSI/ISA-99.02.01-2009* - In 2008 NIST released a comprehensive guidance on securing SCADA systems in the special paper 800-82, Guide to Industrial Control Systems (ICS) Security. This document addresses issues ranging from an outline of SCADA systems, security program development and technical controls and network architecture.
- ITSEAG Generic SCADA Risk Management Framework for Australian Critical Infrastructure [Generic SCADA Risk Management Framework for Australian Critical Infrastructure].
- TISN Critical Infrastructure Resilience Strategy [Critical Infrastructure Resilience Strategy]

In general, ISO standards address issues that cover various aspects of effective management of SCADA systems – including: *organization of information security*; *asset management*; *human resources security*; *physical and environmental security*; *communications and operations management*; *access control*; *information systems acquisition*, *development* & *maintenance*; *information security incident management*; *business continuity management and compliance.*

6. Conclusions

This paper presented a discussion that reflects the body of knowledge characterized as broader strategic security management and governance of complex IT systems such as SCADA systems. The objective of the paper is to raise awareness of importance of security and emphasize that only strategic/holistic approaches (considering key elements such as technology, people and processes) to security management can produce effective outcomes. Based on review of cases and previous studies, this paper outlined broader cyber security issues before introducing SCADA systems. Next, risks and vulnerabilities were investigated before cyber security protection measures were discussed.

Vulnerabilities and tools for attacks were identified to include: Viruses, Trojans & Worms, System & Boot Record Infectors, Eavesdropping, Hacking, IP Spoofing Attacks, Denial of Service, Email Bombing & Spamming, Phishing and Man-in-the-middle attacks. The sources of attacks could fall into any of the following categories: Bot network operators, Terrorists, State Sponsored Hackers, Criminal groups (organized crime hackers), Insider attacks (Disgruntled employees), Hobbyists, Hacktivists and Unskilled key personnel.

The paper suggested that based on the outcome of analysis of risks and threats, a holistic set of strategies must be considered to address various elements that contribute to the security of operation of a system. They are: technology, people and processes. To ensure all those elements and longer term issues are taken into account, a strategic model involving strategic analysis, strategic choice of options and strategy implementation was recommended.

References

Adeyinka O. (2008) Internet Attack Methods and Internet Security Technology. Second Asia International Conference on Modelling & Simulation.

Asgarkhani, M. (2011) A Strategic Outlook for Managing Information and Communication Technology: An Overview. 2011 International Conference on Financial Management and Economics IPEDR vol.11, pp 140-144.

Asgarkhani, M. (2013) Corporate ICT Governance: A Tool for ICT Best Practice. Proceedings of the International Conference on Management, Leadership and Governance. 1: 1-8. 978-1-909507-00-5.

BBC. The Chernobyl disaster (2004) Viewed March 2011. <http://www.bbc.co.uk/dna/h2g2/A2922103>.

Blau J. (2004) The battle against cyber terror. Computer World. Accessed on 10.04.11. URL: http://www.computerworld.com/s/article/ 97953/The_battle_against_cyberterror; December 2004.

Goodin D. Pro-palestine vandals deface army; NATO sites (2008) The Register. Viewed Apr 2011 <URL:http://www.theregister.co.uk/2009/01/10/army_nato_sites_defaced/

Hayes S., Shore M. and Jakeman M. (2012) The Changing Face of Cyber Security. ISACA Journal, volume 6.

Iversen W. (2004) Hackers step up SCADA attacks. Automation World. accessed on 10.04.11. URL: http://www.automationworld.com/ webonly-898;

Mellado D. and Carlos R. (2012) An Overview of Current Information Systems Security - Challenges and Innovations, Journal of Universal Computer Science, vol. 18, no. 12.

Morris T., Vaughn B.R., Sitnikova E. (2013) Advances in the Protection of Critical Infrastructure by Improvement in Industrial Control System Security (ACSW-AISC) 2013, Adelaide, Australia.

Nicholson A., Webber S., Dyer S., Patel T. and Janicke H. (2012) SCADA security in the light of Cyber-Warfare", Journal of computers & Security, 31, 418-436.

Ponemon Institute. The Impact of Cybercrime on Business (2012) Viewed Aug 2013 <http://www.checkpoint.com/products/downloads/whitepapers/ponemon-cybercrime-2012.pdf>.

Robles, R.J., Choi, M. (2009) Assessment of the Vulnerabilities of SCADA, Control Systems and Critical Infrastructure Systems. International Journal of Grid and Distributed Computing Vol.2, No.2.

Sachdeva M., Singh G., Kumar K. and Singh K. (2010) "DDoS Incidents and their Impact: A Review", The International Arab Journal of Information Technology, Vol. 7, No. 1, January

SCADA Honeynet Project. building honeypots for industrial networks. Viewed (2003) Dec 2013 http://scadahoneynet.sourceforge.net/.

Slay, J. & Sitnikova, E. (2009) The Development of a Generic Framework for the Forensic Analysis of SCADA and Process Control Systems. In Forensics in Telecommunications, Information and Multimedia. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, Springer , Vol. 8, pp.76-81.

Susanto H., Almunawar M. and Tuan, Y. (2011) Information Security Management System Standards: A Comparative Study of the Big Five, International Journal of Electrical & Computer Sciences IJECS-IJENS Vol. 11, No 5.

Fuzzy Application With Expert System for Conducting Information Security Risk Analysis

Jiří Bartoš, Bogdan Walek, Cyril Klimeš and Radim Farana
University of Ostrava, Ostrava, Czech Republic
jiri.bartos@osu.cz
bogdan.walek@osu.cz
cyril.klimes@osu.cz
radim.farana@osu.cz

Abstract: This article describes implemented fuzzy tool and extended application for conducting information security risk analysis. The paper describes the fuzzy tool and linked application and shows the advantages of this approach in the contradiction of tools and methodologies used today. This paper covers the fuzzy tool itself and shows its implementation in the process of risk management. The tool is connected to the expert system and methodology which is the part of the complex approach to decision making process that will be demonstrated in this paper. The knowledge base of expert system is created based on user input values and the knowledge of the problem domain. The implemented application is demonstrated on examples and problems from the area of information security.

Keywords: fuzzy, fuzzy tool, expert system, uncertainty, risk analysis, information security

1. Introduction

Today one of the most studied areas is information security and IT/IS environment security. It deals with the protection of people, organizations and their information systems against the dangers of anti-social behavior, physical hazards, technical risk and human errors, rather than risks that are more risks of operational or managerial (business, political, social, economic-financial, etc.). The fundamental concept in this area is the risk analysis (specifically, a risk analysis of information security as a system of analysis of IS/IT environment - information technology and information systems). Risk analysis is a part of the security management, which consists of a series of preceding and subsequent steps that are being cyclically repeated (ISO/IEC 27001:2013). Risk analysis is the most appropriate entry point into the process of managing information security (security of information technology and systems).

Risk analysis is the most appropriate and needed entry point into the process of managing information security (security of information technology and systems). It is also the most serious step and the basis for determination of the security policy and is afterwards often an integral part of risk management processes (ISO/IEC 27002:2013). The risk analysis gives , in the risk management processes, feedback, thanks to which we can evaluate the effectiveness of the implemented measures, the quality of the asset evaluation and of course the whole model of risks.

This article proposes fuzzy tool for conducting risk analysis in the field of information security. This paper is a continuation of article (Walek, Bartoš, Žáček, 2013).

2. Problem formulation

As already mentioned in the introduction, the importance of information security risk analysis is obvious. The problem with conducting the risk analysis correctly lies in the design of appropriate countermeasures to reduce identified risks.

Currently, there are two basic types of methods for risk analysis: qualitative and quantitative (and of course theirs combined approach).

Qualitative methods represent the risk by the elements of a limited set (CARNet, 2003) of typically a numeric interval (e.g. interval <1, 10>, percentage terms, etc.) - individual evaluations are therefore relative and not an absolute numbers (Krutz, 2001). A typical, and probably the best known representative of this approach is method of targeted interviews (Delphi method) based on a set of targeted and specific questions, during which respondents define assets, identify threats and their attributes (Makowski, 2013).

The problem is huge subjectivity during the whole assessment and design of intervals that are being used for the evaluations. This is reflected not only in the quality of the results of the risk analysis but mainly in the fundamental contradiction with the requirement for repeatability of risk analysis is – in other words, once proposed assessment must be used afterwards in every iteration.

Quantitative methods, on the other side, are always based on some mathematical formula for calculating the risk (which is always an exact numerical value), and mutually differ primarily by the complexity of the used formula and the number of variables that are enter that formula.

An example of the calculation formula can be determined by the three factors (Hrvoje, 2006):

$$\text{Risk} = A * H * Z$$

A – asset value
H – the probability of the threat occurrence
Z – vulnerability

In the need for more detailed analysis additional features can be added. It is clear, that the results (risks) generated by the quantitative approach are exact (absolute) values (Šegudović, 2003).

Clean advantages and disadvantages lead to combination of these two approaches into one. These combined methods used semi-qualitative evaluations for the input variables and the calculation formula for the risk evaluation. Overall these combined methods are, based on the analysis phases considered as qualitative.

From the above, some basic differences, advantages and disadvantages between nowadays methods can be deduced:

Table 1: Differences in today's approaches

	Quantitative	**Qualitative**
Advantages	Risks are sorted by financial impacts; assets are classified and identified by their financial value. The results are easy to interpret as a return of investment and objectification of investments in countermeasures. Results can be expressed in terminology understandable for the management of organization (financial value and the probability expressed as a percentage). The accuracy of the analysis is increased over time (thanks to existence of historical data and growth of experience with analyzing of threats).	Allows simple understanding of the computations (evaluation) of risk. It is easier to reach a consensus with partial and total results across all participants involved in the process of risk analysis. It is not necessary to determine the value of the assets (which is time consuming process). It is easier to invite more participants who are not experts in security (but they cannot make the evaluation).
Disadvantages	Values of impacts associated with risks (financial impact) are highly subjective values. The risk analysis process is very time consuming. The risk calculation can be complex and time consuming. The results are presented as financial values and can (often are) difficult to interpret. The analysis process requires expert knowledge and therefore the implementation is problematic. Also, there is a need for the training of all of the participants.	There is a big lack of distinction between major risks. It is very difficult to justify the investment costs; because of the lack of financial evaluation of assets (i.e. there is no basic or trivial correlation). The results depend on the quality of the ream that is performing the analysis.

3. Problem solution

Method of solution is based primarily on the shortcomings that appear in the currently used methods (listed in Section 2). Among other things, the solution is utilizing the variables involved in the evaluation through the use of linguistic variables. In principle, the solution is a combination of qualitative and quantitative (i.e. modification) of today's approaches.

This is done, by the proposal and implementation of fuzzy tool, which is connected to an expert system. The proposed expert system is actually composed from several sub-expert systems to support the whole decision making process. The tool is designed as web application connected to a database system. The application implements an input questionnaire, which is used for asset identification. Afterwards, relevant threats are assigned using set of IF-THEN rules. On the grounds of the selected relevant threats, the risk values are computed and based on the treats and asset types and risk values, relevant and suitable countermeasures are proposed by the system. These countermeasures are visualized in user friendly manner, with attributes describing their effectiveness and implementation price. On the grounds of these attributes the user can decide, whenever to implement them.

Implemented application runs on Apache web server and MySQL database. The application logic utilizes PHP language, HTML5, CSS3 and jQuery framework.

The proposed fuzzy tool with expert system is visually displayed in the following figure:

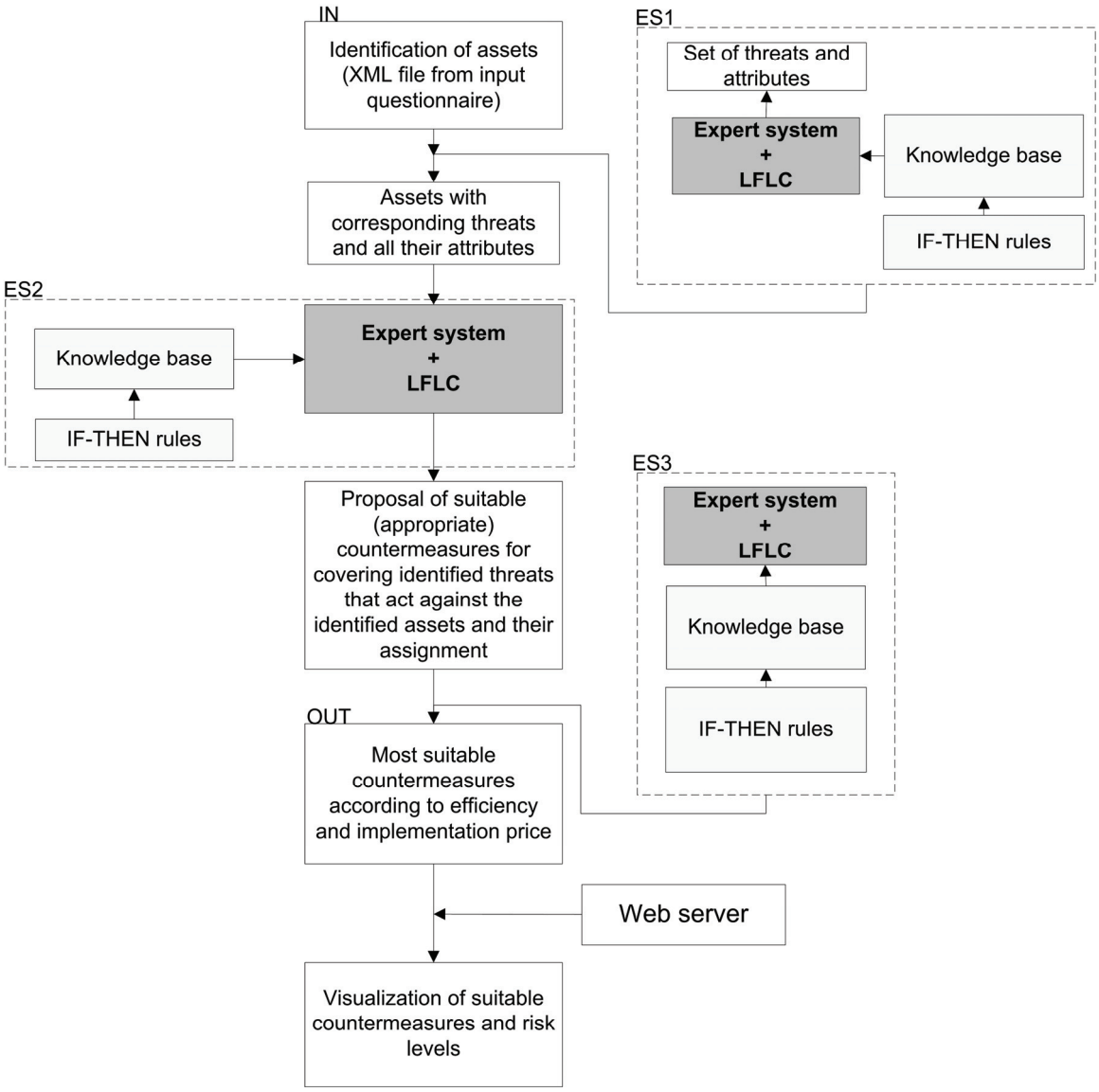
IN
Identification of assets (XML file from input questionnaire)
Assets with corresponding threats and all their attributes
ES1
Set of threats and attributes
Expert system + LFLC
Knowledge base
IF-THEN rules
ES2
Knowledge base
Expert system + LFLC
IF-THEN rules
ES3
Expert system + LFLC
Knowledge base
IF-THEN rules
Proposal of suitable (appropriate) countermeasures for covering identified threats that act against the identified assets and their assignment
OUT
Most suitable countermeasures according to efficiency and implementation price
Web server
Visualization of suitable countermeasures and risk levels

Figure 1: Proposed fuzzy tool with expert system

As mentioned above, the proposed expert system is actually composed from several sub-expert systems and consists of several parts that will be introduced in following sections.

3.1 Identification of assets

In the first step, it is necessary to identify all the important assets. For this purpose an input web questionnaire containing following items is served:

- Asset name – text representation of the name of the identified asset,
- Asset type – selection from the code list,
- Asset value – fuzzy linguistic variable,
- Dependency – fuzzy linguistic variable, defining weight laid on a given asset.

In the questionnaire, the last two items are represented as fuzzy linguistic variables, because the user is often not able to quantify the content of these items exactly. Therefore, the application is suggesting linguistic values, which are closer to the human cogitation.

The questionnaire (part of it), is shown in the following figure:

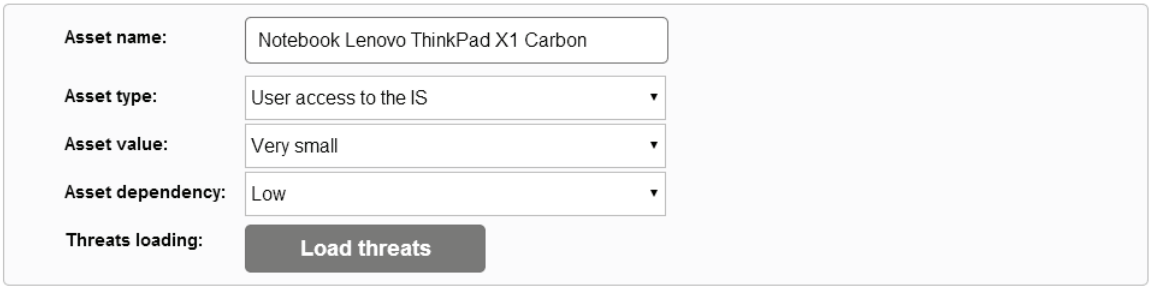

Figure 2: Input questionnaire

3.2 Assignment of relevant threats that is based on the grounds of asset type and value

Based on a filled-in questionnaire, threats will be assigned in the next step. Considering the marked asset type and its value, relevant threats will be selected and assigned. The selection is done by the first set of IF-THEN rules, and utilization of the database of threats (ISO/IEC 27005:2011). This step then leads to completion of the filled-in questionnaire, to its extension respectively. The questionnaire will be extended with the following information:

- Threat name – a threat which can have an impact on a given asset,
- Threat frequency - fuzzy linguistic variable, values: very low, low, medium, high, very high,
- Vulnerability – defines the level of asset vulnerability in the case of threat impact, fuzzy linguistic variable, values: very low, low, medium, high, very high.

This subsystem is described visually in Figure 3:

3.3 Computation of risk values – via the second set of fuzzy IF-THEN rules

This step is closely linked with the previous one. In this step are risks being evaluated – via the second set of fuzzy IF-THEN rules. The knowledge based is created in LFLC tool. LFLC tool is more described in (Habiballa, Novák, Dvořák, Pavliska, 2003). The rules have very simple notation. Output of the expert system is the linguistic variables "risk level":

- Risk level – fuzzy linguistic variable, representing the risk value for individual threat

An example of IF-THEN rules is shown below (this expert system uses 215 IF-THEN rules):

IF asset_value==low AND threat (low, low) AND weight==low THEN risk=acceptable
IF asset_value==low AND threat (low, medium) weight==low THEN risk=acceptable
IF asset_value==low AND threat (low, high) weight==low THEN risk=acceptable
…
IF asset_value==medium AND threat (medium, low) weight==low THEN risk=acceptable
IF asset_value==medium AND threat (medium, medium) AND weight==medium THEN risk=very low

IF asset_value==medium AND threat (medium, medium) AND weight==high THEN risk=high

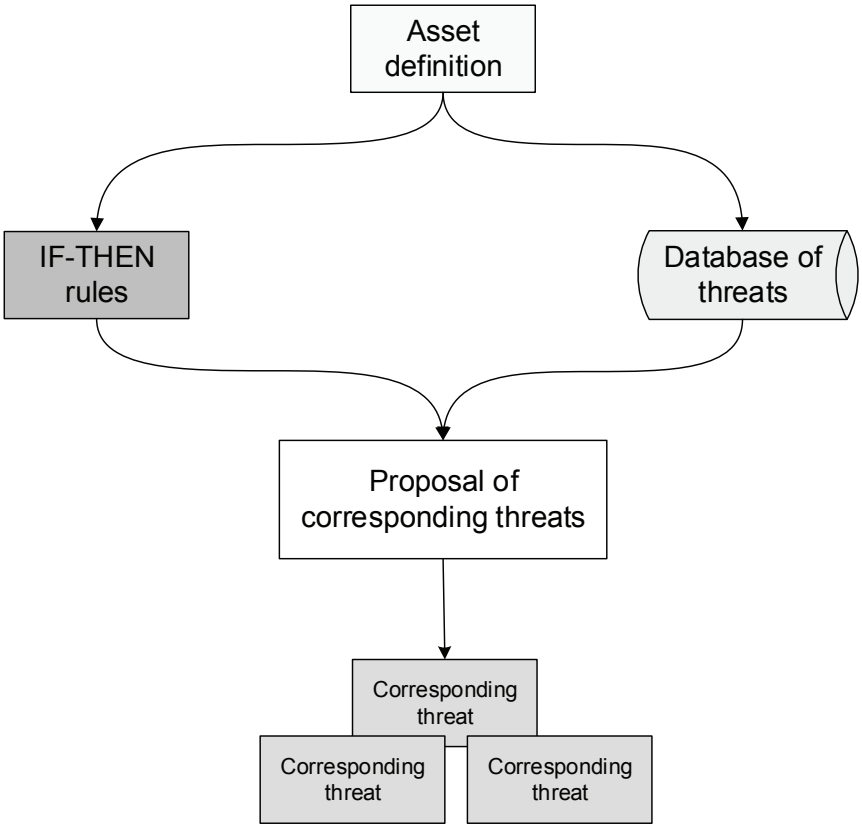

Figure 3: Expert system for assignment of relevant threats

The selection of the relevant threats as presented by the application is illustrated in the following figure:

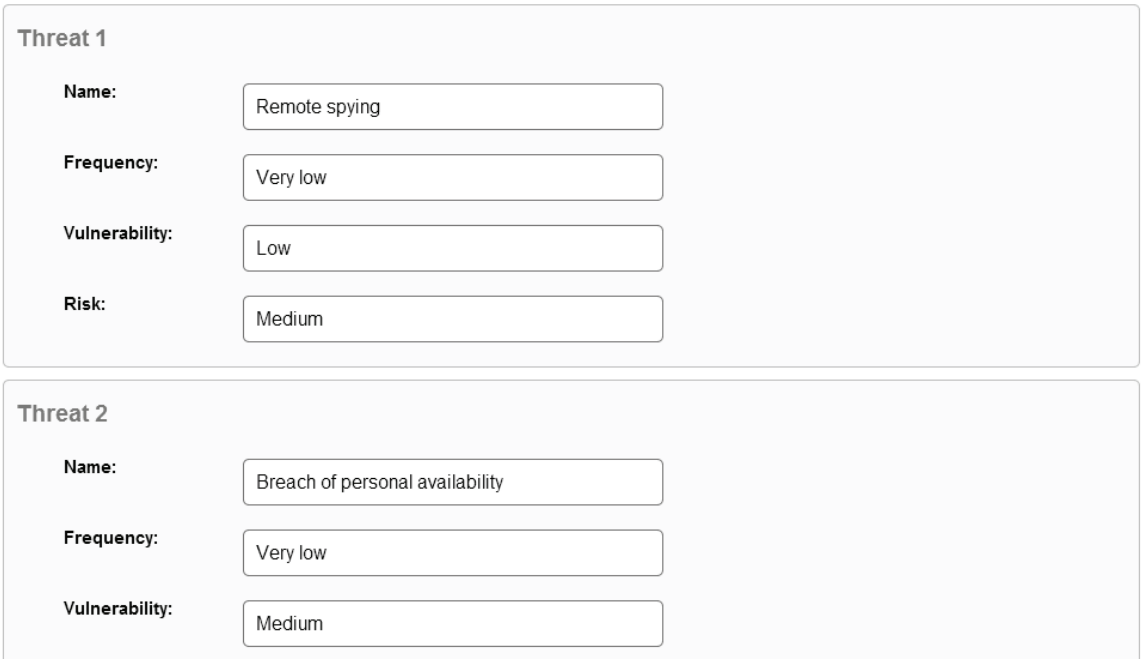

Figure 4: The selection of relevant threats

For every individual threat, there is a risk level value. The overall risk level of the asset is afterwards the supreme of the individual risk level values, therefore the web application loads the supreme, on the grounds of the individual risk levels, and this supreme will be used for the results and outputs that will follow.

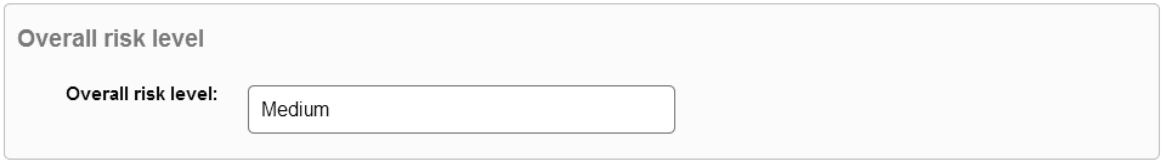

Figure 5: Overall risk level

3.4 Proposal of suitable (appropriate) countermeasures for covering identified threats that act against the identified assets and their assignment – via another set of fuzzy if then rules

Afterwards the most appropriate countermeasures (ISO/IEC 27001:2013) are chosen on the ground of risk levels values, asset types and threats. The knowledge base conducting this selection is again constructed in the LFLC tool. IF-THEN rules for countermeasures selection:

IF asset_value==very low AND risk==acceptable THEN /
IF asset_value== very low AND risk==low THEN corresponding_countermeasures(n,n)
IF asset_value== very low AND risk==medium corresponding_countermeasures(n,n) (n,s)
IF asset_value== very low AND risk==high THEN corresponding_countermeasures(n,n) (n,s) (n,v)

IF asset_value==low AND risk==acceptable THEN corresponding_countermeasures(n,n) (n,s)
IF asset_value==low AND risk==low THEN corresponding_countermeasures(n,n) (n,s)
IF asset_value== low AND risk==medium corresponding_countermeasures(n,n) (n,s) (n,v)
IF asset_value==low AND risk==high THEN corresponding_countermeasures(n,n) (n,s) (n,v) (s,n)

3.5 Visualization of the results of risk assessment

The selection of the most suitable solutions is visualized to do user as a final step of over implementation of the complex expert system. Where, the visualization is done by utilization of modified component model – every component is colored, where every color represents the level of risk evaluation; every subcomponent is colored, where every color represents the appropriate of selected countermeasure:

Component color – risk level:

- Green – low risk
- Yellow – medium risk
- Red – High risk

Subcomponent color – countermeasure effectiveness:

- Green – small efficiency
- Yellow – medium efficiency
- Blue – big efficiency

Figure 6: Visualization of the results of risk assessment

4. Verification

The verification of the tool was conducted on several small and medium size risk analysis projects – because of the extend of the risk analysis; we will use only a snippet of one of them (two assets).

Here are selected assets:

- Asset name: Server room, asset type: Locality

- Asset name: CORE switch, asset type: LAN

Figure 7: Example of asset called server room

For selected assets, the following threats were loaded:

Server room:

- Theft
- Theft
- Dust, corrosion, freezing
- Water damage
- Fire
- Natural event
- Lost of power supply
- Failure of air conditioning or water supply
- Terrorism
- Unauthorised use of equipment by foreign entities
- Unauthorised use of equipment by internal/contracted entities
- Saturation of information system

C0RE switch:

- Disturbance
- Breach of maintainability
- Interception of compromising interference signals
- Remote spying
- Theft
- Theft
- Dust, corrosion, freezing
- Destruction of wiring
- Interception of compromising interference signals
- System software malfunction
- Failure of network devices
- Unauthorised use of equipment by foreign entities
- Unauthorised use of equipment by internal/contracted entities
- Eavesdropping
- Saturation of information system

Then, the overall risk level was loaded:

Server room
Overall risk level - High
CORE switch
Overall risk level – High

Finally, the most appropriate countermeasures were chosen and visualized:

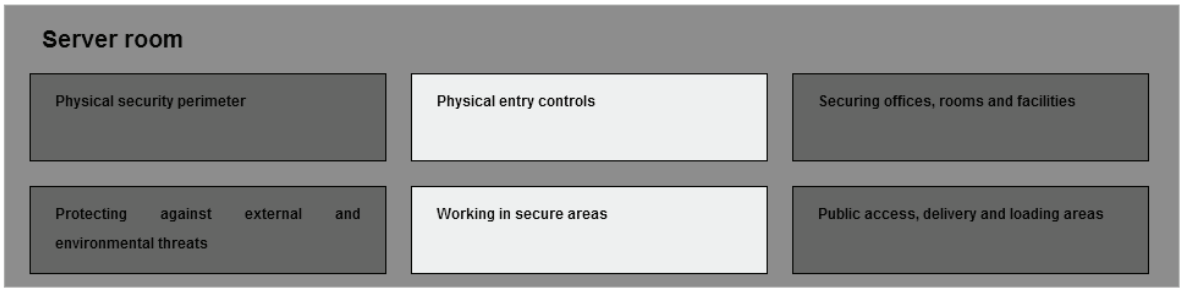
Server room
Physical security perimeter
Physical entry controls
Securing offices, rooms and facilities
Protecting against external and environmental threats
Working in secure areas
Public access, delivery and loading areas

Figure 8: Visualization for asset called server room

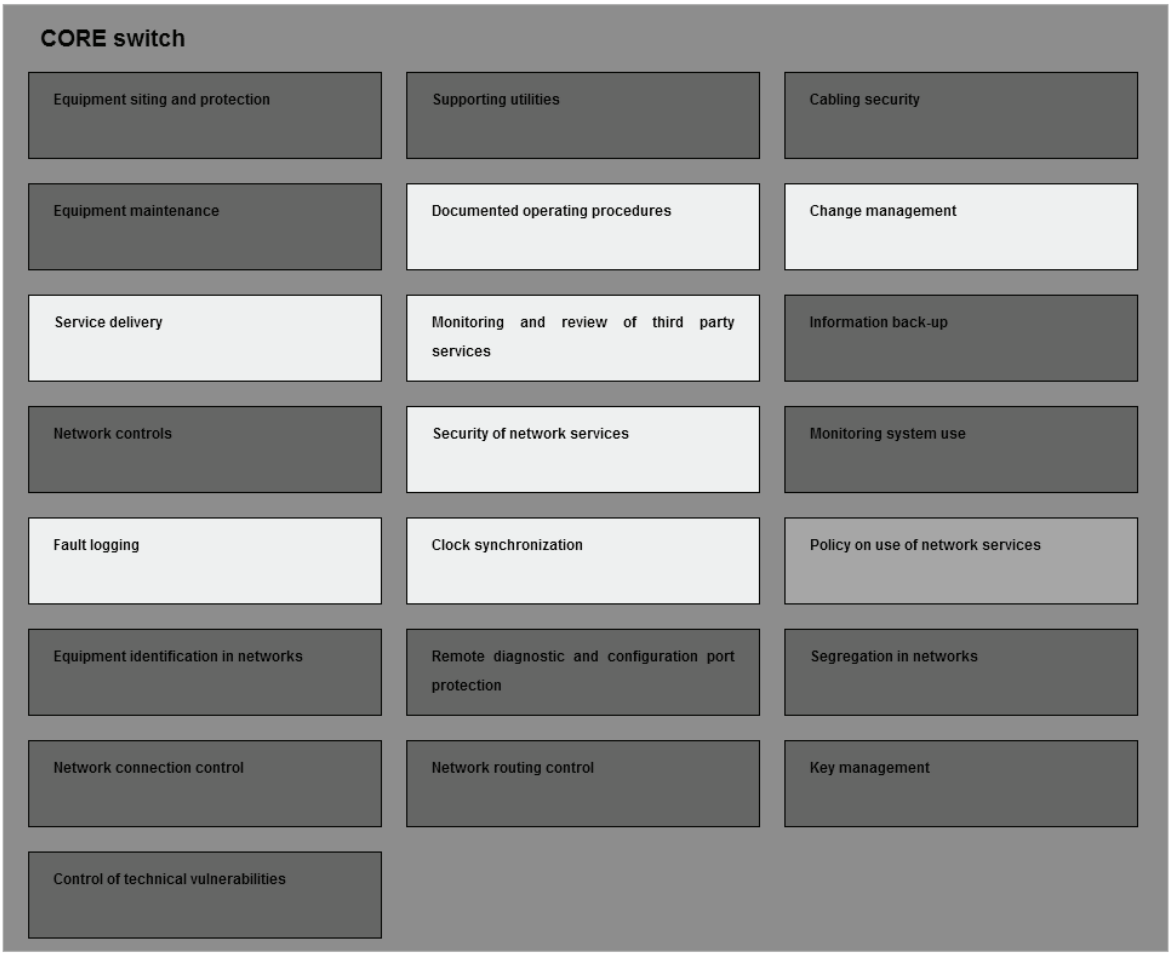

Figure 9: Visualization for asset called CORE switch

5. Conclusion

This article describes implemented application, which is using the fuzzy tool and decision making driven expert system for proposing and selecting appropriate countermeasures to reduce the risks to the assets of the organization and to computes such risks. With expert on security, who already filled all the knowledge bases of the expert systems (basically, every step of the tool in the process of risk analysis utilizes its own set of IF-THEN rules (own knowledgebase) to conduct the decision making), the system is an automated system, without the further need of an expert. The system is already implemented. IF-THEN-rules of knowledge base cooperate

with the goal to minimize all the shortcomings of usual approaches to risk analysis (and therefore can be used not only for risk analysis within the information security area). Since the model uses the knowledge expressed in vague language, the necessary data need not be collected by experts. Moreover, as it was empirically proved in applications, when used by the experts, the risk analysis is conducted in a surprisingly quick, clear, easy and comprehensive way, with the ability to support the understanding of the processes and results to all of the users.

Acknowledgements

The presented topic is also a part of the internal grant SGS15/PřF/2014, called Fuzzy modeling tools for securing intelligent systems and adaptive search burdened with indeterminacy, at the Department of Informatics and Computers, University of Ostrava.

References

ISO/IEC 27001:2013 Information technology - Security techniques - Information security management systems – Requirements, ISO/IEC JTC 1/SC 27, edition 2, 2013.

ISO/IEC 27002:2013 Information technology - Security techniques - Code of practice for information security controls, ISO/IEC JTC 1/SC 27, edition 2, 2013.

ISO/IEC 27005:2011 Information technology — Security techniques — Information security risk management, ISO/IEC JTC 1/SC 27, edition 2, 2011.

Hrvoje Segudivuxm Quantitative risk analysis method comparison, 2006 MIPRO 2007 conference, ISS, pp. 1-6

Krutz R. L. , Vines R. D., The CISSP Prep Guide—Mastering the Ten Domains of Computer Security, John Wiley & ons, Inc., 2001, pp. 12-33.

Šegudović H., Upravljanje sigurnošću informacijskih sustava, KOM 2003, 2003, pp 3-31.

CARNet CERT & LSS, Upravljanje sigurnosnim rizicima, CARNet CERT, 2003, pp. 5-16.

Makowski M., Mathematical Modeling for Coping with Uncertainty and Risk, 2003, pp. 1-20.

Habiballa H., Novák V., Dvořák A., Pavliska V., Using software package LFLC 2000, 2nd International Conference Aplimat 2003, Bratislava, 2003, pp. 355-358.

Walek B., Bartoš J., Žáček J., Proposal of The Expert System for Conducting Information Security Risk Analysis, Proceedings of the International Conference on Electrical and Electronics Engineering, Clean Energy and Green Computing, Dubai, 2013, pp. 58-68.

Identity Multipliers and the Mistaken Twittering of 'Birds of a Feather'

David Cook
Edith Cowan University Security Research Institute, Joondalup, Australia
d.cook@ecu.edu.au

Abstract: New Media usage has expanded in unexpected directions. This qualitative study pulls together two new media phenomena that demonstrate widespread engineered influence through Twitter dispersion. The first aspect is the practice of sock-puppetry deception that occurs when social media identities are used without proof of identity. The relative ease in creating a single invented online identity is overshadowed by the more recent practise of harnessing thousands of created identities to accentuate the standing of a single identity. The second aspect is the use of metadata to fetch and find increased nodes of commonality. This promulgates the perception of connectivity and homogeneity in terms of information transfer. Identity masking in Twitter, and the use of *homophily* to manipulate authority, public sway, and social media deception are a formidable combination. This study considers the increasing use of sock-puppetry in conjunction with techniques that exploit *homophily* to sway the perceptions of others. The ethical and socio-political issues of identity deception and their subsequent impact are discussed. Using the publicly available Twitter accounts of six high profile political figures, the study compares retweets that retain metadata from original tweets. Using both real and fake accounts, the paper tests the hypothesis that non-genuine Sybil Twitter accounts are deployed in the followings of popular figures by using metadata in retweets to establish credibility amongst fake entities. New Media experiences augmented manipulation in multiple ways. The use of metatagging to expose greater commonality increases both the frequency and the amplitude of information diffusion when delivered through social networking and microblogging. A comparison of the associated tweets, followers, and retweets of high profile leaders indicate pervasive sock-puppetry and *Sybil* deception. With *homophily* embedded into Tweets and Retweets, sock puppetry in social media presents a challenge. Where *homophily* fits traditional notions of "similarity between two nodes", increasing the likelihood of likeness to occur is problematic. If *homophily* and false identity coalesce, then the use of twitter as a reliable social medium is contingent on discrete user identity in social media.

Keywords: homophily, Twitter, deception, sock-puppetry, metatagging, Sybil accounts

1. Introduction: The deceptive progression of new media

New Media usage has expanded beyond the expectations of society in general. It has replaced print media, postal mail, and a variety of community-based communications platforms. Twitter in particular, with its high speed delivery and narrative brevity, has bridged the gap between formally reported journalism and the public and civic reporting that has been informally recounted through the community (Jewitt, 2009). Now there is simply news. Individual tweets are now more readily accepted as factual, backed up by smartphone photography, crowd sourced verification, and Twitter's lightning fast diffusion of information (Wilson, 2011). However, as New Media continues its rapid expansion, the volume and frequency of digital telecommunication has accelerated the proliferation of exploits that deceive the world by means of technology force multipliers (Christensen, 2011).

The business community has embraced the marketing value of new media, using the mechanisms of mass connectivity for profit and market share (Wilson, 2009). Users are best described as '*prosumers*', combining content *production* and content *consumption* in a single descriptor (Leavitt, Burchard, Fisher and Gilbert, 2009). Twitter holds prominence within marketing circles because of its ability to connect people with similar thinking, 'like' desires, and, most importantly, analogous decision-making. In business terms this equates to purchases, commitments, profits and revenue (Nielsen, 2009). Such commonalities are tremendously sought after. The business community is repeatedly contemplating whether to use false identities as microblogging entities in order to exaggerate and to substantiate the power and influence of any given message, narrative, or person (Miller, 2009; Stringhini, Wang, Egele, Kruegel, Vigna Zheng, and Zhao, 2013). One of the fundamental golden pillars upon which Twitter is built is the speed and ease of information diffusion. The more people in a social network, and the more connected they are in closeness, the faster and stronger any given message is circulated (Suh, Hong, Pirolli, and Chi, 2010).

1.1 What damage can fake identities do?

Fake identities manipulate, deceive, and grow the uncertainty and mistrust of online communities (Ratkiewicz, Conover, Meiss, Goncalves, Flammini, and Menczer, 2011; Romero, Meeder and Kleinberg, 2011). A fake identity on Facebook or Twitter is most commonly not operating in isolation (Thomas et al., 2012: Ratkiewicz et al, 2011). Whilst sock-puppetry does occur on a personal level expressed in the form of one or two additional extensions of an individual's own discrete identity, it is the practice of botnet-enabled Twitter puppets (Sybils) that can be deployed by the thousand, that is more influential in propagating Twitter identity deception (Krebs, 2011; Wheatley, 2013; Parmelee & Bichard, 2013).

The ethical and legal question about whether the wider *netizenry* should accept the use of multiple invented personas by individuals is most explicitly identifiable in an examination of Twitter entities (Parmelee & Bichard, 2013). The question arises as to whether the public have the right to know when one person's message is deceptively repeated through multiple fake entities to give the impression of widespread support for a person, their theme, or their narrative. Even though the practice of sock-puppetry far outdates the invention of web-based information transfer (Rollins, 1993), the use of technology-enabled puppetry presents a much more powerful and highly augmented set of possibilities in terms of multiple automated microblogging (Chu, Gianvecchio, Wang, & Jajodia, 2010).

The use of such deception is visible and prominent across a wide range of Twitter disciplines, whether it be in business (Streitfield, 2012), violence and extremism (Conway, 2012), or government and politics (Cogburn & Espinoze-Vasquez, 2011). There are already established and institutionalised practices that deliberately augment followings for their own means and ends, whilst at the same time the presence of widespread acknowledgement and recognition of the practices seems largely overshadowed by the more prominent behaviour of ignorance, passive-agnosticism, and tolerance, best generally described under the terminology of *slacktivism*. Fake identities in large enough numbers can dominate or at least influence global outcomes in terms of purchasing behaviours, radicalisation, governance and democracy (Waugh, Abdipanah, Hashemi, Abdul Rahman, and Cook, 2013).

2. New media platforms encourage multiple identities

At a time when individuals are beginning to understand that their data can be used by corporations to great commercial benefit, the notion that social media platforms knowingly continue to promote practices that encourage multiple identities is divisive. Profit, power, influence and authority are drivers of the practice of using botnets and automation to harness the advantages of vastly enlarged support from multiple followings (Stringhini et al, 2013). The Russian parliamentary elections of 2011 saw pro-Kremlin and anti-Kremlin parties endure the deployment of 25860 fake Sybil Twitter accounts used to leverage over 400,000 tweets in a swarm attack that disrupted and confused online narrative *homophily* beyond expectations (Thomas, Grier and Paxson, 2012). The tactic of injecting creative narrative into social networks at key moments reveals one of the major drawbacks of a social media platform that does not require identity authentication beyond the originating CAPTCHA authentication (Twitter, 2013). This would suggest that the organisational side of Twitter is comfortable with the knowledge and expectation that there are many fake Twitter entities, and that large scale sock-puppetry is alive and well in the Twitter community (D'Yonfro, 2013). Twitter does not offer or promote any serious effort towards the filtering of spam (Grier, Thomas, Paxson and Zhang, 2010). Thus, for those seeking to exact influence and authority, Twitter has sufficient leeway and latitude to suggest that multiple identities are accepted as an applicable component of new media.

3. The effect of homophily upon Twitter-based social network influence

The '*Birds of a Feather'* understanding of *homophily* describes the phenomena where communication amongst "similar people occurs at a higher rate than among dissimilar people" (McPherson, Smith-Lovin, and Cook, 2001). Indeed, in most social networks the key pointer to their *homophily* is the reciprocity which denotes their sharing of one or more interests. The reciprocity of social connections shows the exchange and interaction between people and although their reciprocity need not be equally mutual, there is an expectation that the interaction is not all one-way traffic (Weng, Lim, Jiang and He, 2010). However, the same is not true for popular Twitter accounts where a person of celebrity and reputation will typically have many followers but have very few followings (Kwak, Park and Moon, 2010).

Twitter uses *homophily* to draw attention to both the trending themes and ideas of the day as well as those Twitter accounts that command the greatest numbers of followers, and the greatest number of accounts that retweet across those themes and ideas. These trends are readily identified through a range of metatagged expressions that show connection with (although not necessarily support for) a given idea, topic, or narrative expression (Blanquart and Cook, 2013; Waugh et al, 2013). We use the term metatagging here to include the use of identifiable search physiognomies such as hashtags, URLs, attached pics, @ symbols and the acronym RT. These characteristics form the metadata of choice within the Twitter community (Romero, Meeder, and Kleinberg, 2011). In combination with idioms, sarcastic phrasing, and emotive expressions, they shape much of the trend-creating behaviour that is magnetic to twitterbots, cyborgs, and other Twitter manipulators (Chang, 2010; Romero et al, 2011; Wang, Wei, Liu, Zhou, and Zhang, 2011).

The interdependency of entities / actors, and the manner in which their interactions overlap, forms the cornerstone upon which social network influence can be understood (Kwak, Lee, Park, and Moon, 2010). In its most basic sense, the very use of Twitter marks a form of technology acceptance and an approval of the system of microblogging as an appropriate method of dispersing information. Rogers' (2003) work on diffusion is very clearly visible and accurately demonstrated when examining the manner in which an original tweet becomes transferred and interrelated through a network of other retweeted narratives (Chang, 2010). For those in the business of seeking influence and profit, *homophily* is the agent that augments that influence. At the same time, *homophily* should still be understood as the characteristic of individuals who are serious about finding other people (or networks of other people) who share some of their own "sociodemographic, behavioural, and intrapersonal characteristics" (McPherson, Smith-Lovin, and Cook, 2001).

There is evidence to suggest that *homophily* is not quite as innocent a characteristic as often portrayed. In particular, when combined with the inherent multipliers of its metadata, *homophily* and sock-puppetry represent a dangerous combination, in the sway and influence of social media activity (Ratkiewicz et al., 2011; Romero, et al., 2011, and Suh et al., 2010). This combination is an obvious characteristic of our communications such that in normal, everyday terms its effect goes largely unnoticed. Twitter users gravitate towards friends and colleagues assisted by a range of communications that hold them to agree with or share interest in a given person, message, story or outcome. People do not have to like the other person's view, they simply have to be drawn towards the general theme that is of appeal to them. (McPherson et al, 2001; Kwak et al 2010). Metadata has become the catalytic agent of choice for the discerning Twitter entity looking to exert greater influence and authority Ratkiewicz et al., 2011).

3.1 Homophily and influence from actionable content

There are two classifications of tweets on the Twitter spectrum. If a Tweet is 'conversational' then the original tweet and retweet (RT) may respond back and forth as a conversation of sorts ensues. If a tweet (especially from a celebrity or person of prominence) is 'content based' then the direction of the original tweet and any subsequent tweets is more one directional. Subsequent retweets continue to diffuse the content through a wider and wider network. In these situations the original tweet is less likely to give direct responses to replying retweets, and will instead post further 'content-based' tweets as the wider community respond and react. Thus the one directional actionable content is of specific interest in determining the true measure of influence on the Twitter platform.

Leavitt et al, (2009) suggest that there are two extrinsic classifiers that relate to this form of actionable content: the *mention* and the *attribution*.

In the first instance a user may express something like:

Mention: {content} @username ({content})
Watching @BarackObama speak in Colorado on @CNN
RareAir24 (on 2009-08-15 at 19:08:51)

In the second instance the user may narrate as follows:

Attribution: {content} via @username ({content})
Fire at Kuwaiti wedding kills dozens, official media says http://bit.ly/wn95A (via@cnnbrk)

ChilliGaz (on 2009-08-15 at 19:40:18)

Both *mentions* and *attributions* deliver content that when augmented by metadata, (for example hashtag, @CNN, URL, pic) can achieve much higher rates of diffusion through perceived *homophily*. These metadata not only work as organisational markers that assist Twitter in determining its 'trending topics' (Chang, 2010), they also build trust and credibility into narrative attempts at projecting social norms, opinion, and innovation and change (Rogers, 2003). There is widespread evidence to indicate that Twitter *homophily* is augmented by the addition of metadata (Chang, 2010; Blanquart & Cook, 2013; Waugh et al, 2013). This is especially clear where metadata takes the broader meaning to include hashtags, URLs, pics, idioms, RTs, and sarcastic or hateful phrasing (Suh et al, 2010; Kwak et al, 2010; Ratkiewicz et al, 2011; Davidov, Tsur and Rappoport, 2010).

4. Research method

In order to test the followership of high profile political figures for the combinational occurrence of both fake personas and retweets with metadata-augmented homophily, the following sample was taken. The Twitter postings and subsequent retweets were captured for six discrete high profile (trending) Twitter accounts. The data was obtained using Twitter's publicly available API and data was gathered by crawling Twitter in sixty day periods between March 2013 and December 2013. All six accounts were political figures who had posted more than 500 tweets and who had over 50,000 followers. The six accounts were selected at random from a predetermined selection of 24 political figures who met the tweet and follower criteria (Table 1). Their subsequent retweeting entities (active followers) were then examined using the Chu et al. (2010) four way test for fake entities. From these followers a random sample of 10 suspected Sybil entities and 10 genuine entities were selected for each of the six original accounts (Table 2).

Table 1: Political figures from which sample retweets were captured

Political Figures	Username	Tweets	Followers	Following
David Cameron - UK Prime Minister	@David_Cameron	674	579300	331
Geert Wilders - Dutch Parliament Chair Party for Freedom	@geertwilderspvv	1006	311000	0
Tony Abbott - Prime Minister of Australia	@TonyAbbottMHR	1570	278000	31500
Ed Milliband - UK Leader of Labour Party	@Ed_Miliband	3000	287600	1446
Bill Shorten - Leader of the Australian Labor Party	@billshortenmp	1162	51340	10600
Dr Manmohan Singh - Prime Minister of India	@PMOIndia	4206	1068317	48

A total of 120 retweeting entities (60 genuine and 60 Sybil) were examined for signs of metadata used as actionable content. A total of 6,895 retweets drawn from all 120 entities were reviewed for their use of like-metadata. In cases where a retweet used metadata that was different from the metadata in the original tweets, the results did not record the retweet as showing homophily-type metadata. For example if a different hashtag was used in a retweet, then the metadata component was not recorded even though the retweet was included in the sample. Only retweets derived from one of the six chosen figures were considered in the sample data. Retweets were considered on the basis of re-stated metadata (notably the use of URL, hashtags, Pics, Youtube links, @username, and phrases constituting sarcasm, hateful commentary and idioms). A comparison was then drawn between the genuine and Sybil accounts in order to see whether non-genuine accounts showed greater levels of homophily-generating metadata.

Table 2: Chu et al. (2010) 4 way test to distinguish genuine from Sybil accounts

	Four Way Test for non-genuine Twitter entities.
1.	*Entropy Test - Measure Retweet intervals*
2.	*Spam and Miscreant Test - Check for Benign or Malicious content*
3.	*Account Properties – Does the Account have subscriber details or does it look hollow*
4.	*Discrimination Analysis – combining Entropy, Spam, and Account Properties to evaluate all three indicators*

5. Results

A comparison between the sixty genuine followers and the sixty non-genuine (Sybil) accounts showed a number of areas where retweeted narrative shared similar behaviour. Since the behaviour of the non-genuine accounts was assumed to use some form of bot-like automation, it was postulated that Twitter-bots, cyborgs and automated Sybil accounts would show equal or greater numbers of discrete retweets that also incorporated one or more of the metadata markers as described above. The results were highly consistent with this hypothesis in terms of URLs, hashtags, and pics. In these three areas the Sybil accounts showed significantly greater numbers of discrete retweets where the metadata was also passed on as part of the retweeted message. The more often a retweet had one or more of these metadata, the more likely it was to be retweeted forwards. Across both genuine and Sybil entities, the strongest indicator of homophily metadata was shown as links to Pics and Youtube clips. Pics and Youtube links in particular showed the highest levels of *homophily* in fake (Sybil) accounts. Since these characteristics are ascribed to fake accounts, it is the perceived commonalities that we describe as *homophily*, rather than actual like-mindedness, since the retweets in question represent contrived puppetry. Nevertheless, these characteristics are indeed consistent with *homophily* as expressed through higher numbers of retweeted actionable content, regardless of whether they are *mentions* or *attributions*.

In contrast the @username metadata showed no discernable differences between fake and real retweets. This would suggest that for the most part, botnets are not relying on the @username metadata to extend or augment the influence and authority of any given Twitter narrative. This is consistent with marketing and business research (Suh et al, 2010) that places greater emphasis upon links to websites, pictures and youtube clips.

6. Discussion

Whilst the retweeting of @username was present in both genuine and Sybil accounts, its presence was lower than that of the URL based objects such as links to Pics, Youtube and an assortment of extended blogs and web pages. This presents an intriguing comparison. The notion of *homophily* is defined as homogeneity to intrapersonal characteristics (McPherson et al. 2001; Weng, et al. 2010). Therefore the @username might be expected to retain a greater overall prominence in retweets. However it is the subject-based URLs rather than the people-based @username that carries the lion's share of the homogenous linkages.

This study also examined descriptive phrasing and narrative in terms of its metadata potential. Phrases, the use of sarcasm, and idioms were all included in the analysis of retweets. These characteristics had a much lower percentage of *homophily* with fake retweet accounts. We suggest that this is because phrases and idioms require individual and personal scrutiny and do not present the same ease of identification as the more obvious and object-oriented metadata such as URLs, Pic links and hashtags. In fact, sarcastic idioms and hateful phrases can become singleton narratives, since they often hold a level of intensity that does not maintain the same fervor through subsequent retweeting entities. Some phrases, particularly where explicit or passionate phrases are narrated, are highly personal and therefore may encourage restraint rather than connection. Those retweets that showed idiom and sarcastic phrasing were usually re-messaged because they also contained hashtags or other metadata as well. In the sample that was collected, there were no significant numbers (less than 0.1%) of tweets where idiom of sarcasm was retweeted unless other metadata was also present in the original tweet.

The identification of non-genuine Twitter entities was determined using the Chu et al (2010) four way test (Table 2). The test identifies automation in bot crawlers through entropy, the timing of retweets and their relevant intervals. It also looks at the spam and miscreancy behaviour of entities, as well as the account details, its stature and potential hollowness. In combination, these patterns show clearly visible differences between Twitter-bots and genuine human Twitter accounts. Since the presence of Twitter-bots and Sybil accounts was predicted as an underlying assumption of the study, the relative ease with which they were identified suggests their widespread integration into political audiences of the Twitter community.

Table 3: Percentage of retweets from sample that deployed one or more pieces of the same metadata expressed by an original tweets as nominated in Table 1

Retweeting Entity Classification Real or Sybil	URL %	Hashtags %	Pics & Youtube %	@username %	Sarcasm / Hateful / Idiom %
Real RTs for Cameron	22.4%	21.2%	39.0%	17.9%	5.6%
Sybil RTs for Cameron	47.4%	34.3%	68.9%	18.5%	4.6%
Real RTs for Wilders	22.3	17.0%	42.8%	7.6%	13.9%
Sybil RTs for Wilders	37.8	25.1%	58.4%	7.6%	5.8%
Real RTs for Abbott	23.3%	22.0%	39.3%	15.6%	4.9%
Sybil RTs for Abbott	36.5%	36.4%	74.2%	15.8%	3.3%
Real RTs for Miliband	19.5%	16.2%	34.1%	11.1%	3.7%
Sybil RTs for Miliband	35.2%	33.0%	53.3%	10.4%	3.7%
Real RTs for Shorten	9.3%	2.3%	42.9%	13.9%	1.6%
Sybil RTs for Shorten	9.1%	3.3%	43.2%	14.3%	1.6%
Real RTs for Singh	3.1%	4.9%	46.3%	2.1%	5.6%
Sybil RTs for Singh	3.3%	3.7%	63.9%	1.9%	4.4%

Note: This table shows the percentage of retweets that included a meta-marker from an original tweet. It does not include those instances where a retweet has included a fresh item of metadata. Thus it is possible that the figures for both real and Sybil retweets may have a slightly higher total number of tweets that include some form or otherwise homophily-like metadata.

6.1 Hypotheses results

This study's hypothesis asked whether large numbers of fake, automated retweets were establishing a false credibility by including specific objects of metadata in retweeted messages. Are Twitter force multipliers in the form of metadata are a means to promulgate influence? If sock-puppetry and *homophily* could be combined, it would be possible for politically driven groups and associations to exercise significantly greater influence upon mass audiences. Thus by deploying large numbers of Sybil accounts that carry greater authority and trustworthiness by establishing desirable homogeneity between the retweeting entities and their original high-trending counterparts, a narrative of sway can be significantly amplified. It is clear from the results that URLs that link to pictures, youtube clips, and other extended blogs form the type of *homophily* that can both augment the credibility of a fake twitter account, and at the same time mask its hollowness.

7. Conclusion

Many researchers have attempted to understand the exact metrics that correctly indicate the size and speed of the influence that can be attributed to each twitter narrative, and subsequently to broader themes and Twitter figures of prominence. To date there are many algorithms but none that accurately and predictably measure genuine influence (Weng et al, 2010). However the presence of Sybil accounts dilutes the value of most metrics which include the growing number of puppet-controlled Twitter accounts. The combination of sock-puppetry and homophily-attributed metadata challenges the notion that "birds of a feather flock together". Twitter influence is at risk of great deception through the disguised and manipulated persuasion using force-multiplied objects of metadata. The descriptor *homophily* has its historical origins in the homogeneity of people and their likes, yet in the Twitter network its power lies more closely with the eye-catching gimmickry of shared pictures and videos.

References

Blanquart, G., Cook, D., (2013), Twitter Influence and Cumulative Perceptions of Extremist Support: A Case Study of Geert Wilders. *The Proceedings of the 4th Australian Counter Terrorism Conference*, 1 - 11, Perth, Western Australia.

Chang, H. (2010). A New Perspective on Twitter Hashtag Use: Diffusion of Innovation Theory. *Proceedings of the American Society of Information Science and Technology*, Volume 47, Issue 1, pp1-4.

Christensen, C. (2011). Twitter Revolutions? Addressing Social Media and Dissent, *The Communication Review*, Volume 14, Issue 3, **DOI:**10.1080/10714421.2011.597235.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on Twitter: Human, Bot, or Cyborg? in the *Proceedings of the 26th Annual Computer Security Applications Conference,* ACSAC '10, pp21-30.

Cogburn & Espinoza-Vasquez, (2011). From networked nominee to networked nation: Examining the impact of web 2.0 and social media on political participation and civic engagement in the 2008 Obama campaign. *Journal of Political Marketing.* Volume 10, pp 189-213.

Conway, M. (2012). *From al-Zarqawi to al-Awlaki: The Emergence of the Internet as a New Form of Violent Radical Milieu*. Retrieved from http://www.isodarco.it/

D'Yonfro, J. (2013) Twitter Admits 5% of its 'users' are Fake. *Business Insider Australia*, Retrieved October 19th from: http://www.businessinsider.com.au/5-of-twitter-monthly-active-users-are-fake-2013-10

Davidov, D., Tsur, O., and Rappoport, A. (2010). Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning,* pp107-116, Uppsala, Sweden.

Grier, C., Thomas, K., Paxson, V., and Zhang, M. (2010). @spam: The Underground on 140 Characters or Less. *Proceedings of the 17th ACM conference on Computer and Communications Security*, pp27-37, DOI: 10.1145/1866307.186311

Jewitt, R., (2009). Commentaries: The trouble with twittering: Integrating social media into mainstream news. *International Journal of Media and Cultural Politics,* **5***(*3*),* 233–246. DOI: 10.1386/macp.5.3.233/3

Krebs, B. (2011). Twitter Bots Drown out Anti-Kremlin Tweets, *Krebs on Security*: In-depth security news and investigation, Retrieved October 14th 2013 from: https://krebsonsecurity.com/2011/12/twitter-bots-drown-out-anti-kremlin-tweets/

Kwak Lee, C.*,* Park, H.*, &* Moon, S. *(*2010*). What is Twitter: A social network or a news media.* Proceedings of the 19 th International Conference on the World Wide Web (pp. *591–600*). New York, NY: ACM.

Leavit, A., Burchard, E., Fisher, D., and Gilbert, S. (2009) The Influentials: New Approaches for analyzing influence on twitter. *Web Ecology Project*, Retrieved 21st October from: http://tinyurl. com/lzjlzq

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in social networks. *Annual Review of Sociology*, Volume 27 (1) pp 415 – 444.

Miller, C. C. (2009). The Obsession with Twitter's Business Model, *The New York Times*, March 26th, 2009.

Nielsen, (2009) Global Advertising Consumers Trust Real friends and Virtual Strangers the Most, in *Consumer*, Neilsen Global Online Consumer Survey, retrieved 30th October 2013 from: http://www.nielsen.com/us/en/newswire/2009/global-advertising-consumers-trust-real-friends-and-virtual-strangers-the-most.html

Parmelee, J. H. & Bichard, S. L. (2013). *Politics and the Twitter Revolution.* Lanham, Maryland: Lexington Books.

Ratkiewicz, J., Conover, M. D., Meiss, M., Goncalves, B., Flammini, A., and Menczer, F. (2011). Detecting and Tracking Polictical Abuse in Social media. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, retrieved October 29th from: http://truthy.indiana.edu/site_media/pdfs/ratkiewicz_icwsm2011_truthy.pdf

Rogers, E. M. (2003). Diffusion of Innovations, The Free Press, New York

Rollins, D. (1993) Arty/Scotto bit list server Dana Rollins, retrieved 4th October 2013 from: https://groups.google.com/forum/#!msg/bit.listserv.fnord-l/D_wlg9YXFA0/n0aWLZwctdEJ

Streitfield, D. (2012). The best Book Reviews money Can Buy, *The New York Times*, retrieved October 1st 2013 from http://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html?_r=5&pagewanted=all&

Stringhini, G., Wang, G., Egele, M., Kruegel, C., Vigna, G., Zheng, H., and Zhao, B. Y. (2013). Follow the Green: Growth and Dynamics in Twitter Follower Markets. *Proceedings of the 2013 conference on Internet measurement*, pp163-176. DOI: 10.1145/2504730.2504731.

Suh, B., Hong, L., Pirolli, P., and Chi, E.H. (2010). Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. *Proceedings of the 2010 IEEE Second International Conference on Social Computing*, pp177-184.

Thomas, K., Grier, C., and Paxson, V. (2012). Adapting social Spam Infrastructure for Political Censorship, Proceedings of the 5th USENIX conference on LEET; retrieved October 13 2013 from: https://www.usenix.org/system/files/conference/leet12/leet12-final13_0.pdf

Twitter. (2013). *Twitter; An Information Network*, retrieved 13th October 2013 from: https://twitter.com/about

Wang, X., Wei, F., Liu, X., Zhou, M., and Zhang, M. (2011) Topic sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach, *Proceedings of the 20th ACM International Conference on Information and knowledge management.* pp1031-1040. DOI:10.1145/2063576.2063726

Waugh, B., Abdipanah, M., Hashemi, O., Abdul Rahman, S., Cook, D**.,** (2013), The influence and deception of Twitter: the authenticity of the narrative and slacktivism in the Australian electoral process. *The Proceedings of the 14th Australian Information Warfare Conference*, pp28 - 38, Edith Cowan University, Perth Western Australia.

Weng, J., Lim, E, P., Jiang, J. and He, Q. (2010). Twitterrank: finding topic sensitive influential twitters. *In the Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010.

Wheatley, M. (2013). Twitter Shoots down TweetAdder in War on Spam, *Silicon Angle*, retrieved 14th October 2013 from: http://siliconangle.com/blog/2013/05/29/twitter-shoots-down-tweetadder-in-war-on-spam/

Wilson, J. (2009). Social networking: the business case, in *Engineering and Technology*, Volume 4, Issue 10, pp54-56

Wilson, J. (2011). Playing with Politics: Political Fans and Twitter faking in post-broadcast democracy, *Convergence: The International Journal of Research into New Media technologies*. DOI: 10.1177//1354856511414348

Wozniak, G. D. (1987). Human Capital, Information, and the Early Adoption of New Technology, *Journal of Human Resources*, Volume 22, Issue 1, pp 101 -112

Secret Sharing Framework Based on Digital Certificates

Paul Crocker[1,2] and Adolfo Peixinho[1]
[1]Department of Computer Science, University of Beira Interior – 6201-001 – Covilhã, Portugal
[2]Institute of Telecommunications, Covilhã, Portugal
crocker@di.ubi.pt ,
m4067@ubi.pt

Abstract: This paper describes a scheme for using secret sharing cryptographic protocols for sharing, distributing and storing secret keys and documents based on the use of digital certificates. The advantages and uses of such a system in the context of collaborative environments are discussed. A *Framework* for using *Secret Sharing* and in particular, *Threshold Cryptographic Methods*, in order to distribute and share documents (PDF's) and for the granting of authorization and access rights is presented. The paper will describe the cryptographic protocols involved and the necessary Public Key Infrastructure (PKI). An implementation of this *Threshold Cryptography Framework* using an open source PKI Certification Authority is described that uses a *certificate extension* as the basis for storing and accessing the shares. The *Framework* enables users to perform group based decisions based on previously agreed *thresholds* which take into account the *privacy* and *confidentiality* of the entities involved, hence allowing users to manage share and recover authorization *tokens* and documents and manage trust relations in group situations. The framework addresses the key distribution problem as the keys are managed and distributed by the Certificate Authority of the PKI and also deals with hierarchy as it's supported by a PKI, which is inherently hierarchical, hence the different privileges of a user associated with a certificate can be dealt with by the CA. Examples of the *Framework* use are illustrated and the problems and challenges remaining presented. The paper concludes with a practical security assessment of the prototype.

Keywords: secret sharing, digital certificates, certificate extensions

1. Introduction

This paper proposes the implementation of a *Secret Sharing Scheme* (SSS) based on the use of digital certificates. In particular the aim is to define a type of group certificate that can be used by a subset of the group to gain authentication and or access rights for some high value secret objects, such as shared documents and authorization tokens and keys. In this scheme the public certificate of a particular user of the system contains his or her public keys, signed by the certificate authority of the global system. The public keys can be used for strong authentication on the proposed frameworks web sites and the signing of cryptographic hashes, however the main contribution of this paper is to develop a scheme that also permits the digital certificate to contain secret shares thereby providing the user with a single token that can be used to interact with the system web applications and also contain the users secret shares.

The basic enabling technologies for the scheme proposed are *threshold cryptography* and *digital certificate extensions*. A threshold crypto-system, or more simply a secret sharing system, is a system where a secret is divided into a set of shares which are then distributed amongst a set of participants where it is necessary that a number of entities larger than a *"threshold"* value cooperate during the decryption protocol. For instance in sensitive and critical environments in order to increase trust in the system and improve confidentiality and for sheer necessity sensitive keys can be broken up and stored at several physical locations or on several digital certificates. Also to increase security it is often necessary that a subset or quorum of the key-holders or access right holders must simultaneously authenticate themselves in order to gain the *access right* or *secret*. Businesses, Hospitals and even Research departments often need to share sensitive documents and the protection, integrity and confidentiality of these documents is often crucial, cryptographic support for such activities is therefore desirable.

A typical example of *threshold cryptography* is for sharing the secret key of a public key generating system or cloud system deduplication. Another example is where documents need to be stored and shared, for instance consider the situation where two research departments need to share research documents in order to coordinate or authorize research grants or research budgets, etc, but also need to assure that the documents are kept confidential and the power to share these documents needs to be *"limited"* in some way. Consider the example where to access and sign a company document a quorum of four of the seven company directors must be present, the previous scenario denotes a *(4,7) Threshold Scheme* which helps to solve the previously

stated example. All these examples require a *group based decision*, in this case supported by a Threshold Scheme. Threshold Schemes can be used in HSM modules for fault tolerance and for the security of the server keys, an example is SafeNet (SafeNet 2011) Hardware Security Modules (HSMs) which incorporates "M of N" authentication in order to decrypt sensitive keys. However the use of Threshold Schemes and their implementation in group protocols using group certificates is still relatively unexplored.

The rest of this paper is organized as follows: In section 2 we give a brief overview of secret sharing and the secret sharing scheme used in our framework. In section 3 we discuss Public Key Infrastructure, and in particular define how we embed secret shares onto digital certificates. Section 4 describes the architecture and proposed and configurations necessary in order to implement our framework. Section 5 illustrates the implementation of our framework, a case study for sharing PDF documents and gives a security assessment of the implementation. Finally in section 6 we give the conclusions of this paper.

2. Secret sharing

2.1 Brief overview of secret sharing schemes (SSS) and threshold schemes

Secret Sharing Schemes were announced independently by Shamir (Shamir 1979) and Blakley (G. Blakley 1979) in 1979. A SSS begins with a *secret* and derives from it certain *shares* which are distributed to the participants. This *secret* may be recovered only by certain predetermined groups of participants; the group size is the threshold. The set of all groups, or *"qualified subset"*, which can reconstruct the *secret,* is referred to in the literature as *access structure* see for instance (Ito et al. 1987)*.* The reconstruction of the secret can be made by the participants after they pool together their shares or by a special third party, called *combiner* or *dealer,* after receiving the shares from the participants of an *authorized group.* Depending on the *"quantity"* of the secret-information leaked to an unauthorized group, we have the following categories:

- Perfect Secret Sharing Schemes - The shares of any unauthorized group give no information.
- Computationally Secure Secret Sharing Schemes - Some information about the secret is leaked to the unauthorized groups, but the problem of finding the secret is intractable.

From the *Theory of Information* if the length of every share is the same as the length of the secret then these schemes are considered *ideal* (Karnin 1983), Shamir's SSS is such an example. The original SSS by Shamir and Blakley are *(k, n)* threshold SSS. However, there exists variations such as the Ramp Secret Sharing Schemes (G. a. Blakley 1984) (Yamamoto 1986) denominated (k, L, n) threshold Ramp SSS. The (k, L, n) threshold Ramp SSS are designed such that a secret S can be recovered from arbitrary k out of n shares but no information of S can be obtained from arbitrary $(k - L)$ or less shares.

In the case of (k, n) threshold access structure schemes, one normally assumes that every share is equally important, but there are cases we want to make some shares more important than the others, or *hierarchize* the shares. Shamir (Shamir79) proposed a simple modification of the threshold scheme to be used by organizations where every participant receives as its share a certain number of shares from a threshold scheme, according to its position in the hierarchy. In this way a scheme for a *weighted threshold access structure* is obtained; every participant has a *weight* (a positive integer) and a set is qualified if and only if its *weight sum* is at least a given threshold. This scheme is not *ideal* because the shares have in general larger length than the secret.

2.2 Shamir's secret sharing scheme

In this section we shall describe the secret sharing scheme (SSS) as enunciated by Shamir in 1979. The scheme is based on polynomial interpolation and the fact that given k distinct points in a 2-dimensional plane $(x_i, y_i) : i = 1, 2, ..., k$ there is one and only one polynomial $q(x)$ of degree $(k-1)$ such that $\forall i, q(x_i) = y_i$

2.2.1 Distribution of the secret shares

A *Trusted third-party (TTP)* or dealer distributes shares of a secret S to n users where the threshold is defined as the value k.

The TTP sets the secret S as the first coefficient of a random polynomial Q of degree $(k-1)$ over the field of positive integers modulo a large prime p .

The *TTP* selects $(k-1)$ random, independent coefficients $\{a_i\} : i = 1, \cdots, k-1$, defining the random polynomial over $\mathbf{Z}_p : q(x) = a_0 + a_1 x + \ldots + a_{k-1} x^{k-1}$ and sets $a_0 = S$ (the secret).

The *TTP* computes the n shares $S_i = Q(i) \bmod p, 1 < i \le n$ $S_i = Q(i) \bmod p,\ 1 < i \le n$, or for any i distinct points, and securely transfers the share S_i to user U_i , along with public index/point i .

2.2.2 Recovery of secret

Any group of k or more users pool their shares which provide k distinct points $(x, y) = (i, S_i)$ allowing computation of the coefficients of $q(x)$ by Lagrange interpolation, $q(x) = \sum_{i=1}^{k} f(x_k) L_i(x)$ where $L_i(x) = \prod_{j=1, j \neq i}^{k} \frac{(x - x_j)}{(x_i - x_j)}$ and then evaluate $S = Q(0) = \sum_{i=1}^{k} y_i \prod_{j=1, j \neq i}^{k} \frac{-x_j}{x_i - x_j}$ to recover the *secret*.

2.3 Secret sharing software

There are several existing software packages available for secret sharing, these normally consist of a desktop or web application where a single user enters a secret and then the dealer distributes all the shares to the user who is responsible for the storage of the secret shares. An example of such software can be found for example at *http://point-at-infinity.org/ssss.* Another related software is Nightingale from RSA Security which uses secret splitting to secure sensitive data (Brainard 2003) by dividing a user's password (or other key) into shares for two independent servers, the scheme is used in password authentication. (Crocker 2013) describes an implementation of a collaborative mechanism to share secrets between a group of mutually authenticated users based on a chat room with smart card authentication and secret share storage on the smart card. However a generally accepted easy to use and setup for general and business usage is a difficult proposition, for instance smart cards are expensive, difficult to use and often have limited writable file space.

3. Public key infrastructure and X.509 extensions

3.1 Public key infrastructure and X.509 certificates

The secret sharing framework proposed is based on X505 digital certificates and associated public key infrastructure. A *digital certificate* is a signed assertion about a public key. More specifically, it is a *digital signature* of the Certificate Authority (CA) that binds a public key to some other piece of information. This enables all system participants to verify the name-key binding of any presented certificate by applying the public key of the CA. A *public key infrastructure* (PKI) is an infrastructure for a distributed environment that centres on the distribution and management of public keys and digital certificates. It is widely recognized that PKIs are an essential ingredient for secure electronic communications and transactions in open environments. The CA can be made responsible not only for certifying public keys and authenticating certificate applicants, but also for notarizing electronic documents, resolving disputes, and keeping track of revoked keys.

Some or all of these functions may be managed by separate trusted parties. For instance, the registration and approval of certificate applicants may be done by a separate Registration Authority (RA). In practice, a PKI can have multiple CA's, so that certificate applicants and verifiers need not trust a single CA. CA's can certify the public keys of other CA's, and in this manner arbitrary CA structures can be formed. This gives rise to such notions as certificate chains, bridge CA's, and cross certification. These techniques enable anyone to be the issuer of their own digital certificates, and all issuers can coexist in a single PKI.

The X.509 certificate framework is the best known example of *identity certificates.* In 1988, the International Telecommunications Union (ITU) started working on X.509. X.509v1 was designed to certify the public keys of

principals that are uniquely named in X.500 (CCITT 1988), an online database listing globally unique names. An entry in an X.500 directory can be a person, a device, or anything else that can be assigned a "Distinguished Name". The third version of the X.509 certificate framework, X.509v3, released in June 1997, greatly improved the flexibility of X.509 certificates, by providing for a generic mechanism to extend certificates that uses the concept of Object identifier (OID)

An Object Identifier (OID) is an identifier used to name an object (as implied). OIDs are used extensively in computer security. OIDs serve to name almost every object type in X.509 certificates e.g. Distinguished Names, Certificate Policy Statements, Certificate extensions, etc. Structurally, an OID consists of a node in a hierarchically-assigned namespace, formally defined using the ITU-T's ASN.1 standard, X.690. Successive numbers of the nodes, starting at the root of the tree, identify each node in the tree. Users set up new nodes by registering them under the node's registration authority. The root of the tree contains the following three arcs (I) ITU-T (II) ISO and (III) joint-ISO-ITU-T. All other arcs branch from these three.

3.2 Embedding secret shares on digital certificates

Our scheme proposes the embedding of secret sharing shares inside digital certificates as certificate extensions. This solution solves the problem of how to store a user's shares as well as providing the user with a token that can be used to perform strong authentication on the frameworks web services and sites. The use of digital certificates is widespread and varied, from authenticating a users identity on a web site to performing digital signatures. As stated previously our objective is to provide a CA that embeds a security token in the issued digital certificates to support Threshold Cryptography, in this case Shamir's SSS. The custom certificate extension enables this.

The certificate extension has the following requirements:

- A unique specified OID of the extension;
- Full class name (classpath) of the certificate extension implementation class;
- Specification if the certificate extension should be used or be disabled;
- Specification if the display name should be translated in the language resources;
- Specification if the extension should be marked as critical in the certificate;
- Specification if the extension encoding value and if this value is set dynamically or not.

This certificate extension will hold, specifically, 3 values:

- A group identifier;
- The user share for that group, in plain-text;
- A "salted" hashed value of the user share for that group.

Both of these values, also use string terminator symbols, for posterior parsing of the DERUTF8String. Group identifier and user share uses a terminator string symbol of ':' and the *"salted" hashed* user share uses the terminator string symbol of ';'.

In this way the certificate extension can hold several user shares, for several groups, we need a way to know where a user share ends and another begins. The following Figure 1 illustrates an example of such certificate extension.

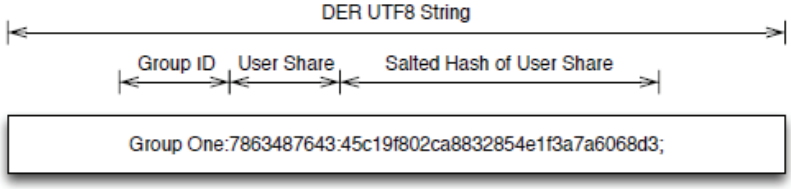

Figure 1: Example of the secret-sharing certificate extension

3.3 Implementing certificate extensions on PKI software

The PKI CA software used to implement our framework is EJBCA from Primekey (EJBCA 2013). This software is a full featured open source enterprise class PKI Certificate Authority built on Java Enterprise Edition technology that is used by numerous high profile organizations and supports the technologies needed to implement our framework, such as X.509 certificate extensions. EJBCA provides a web interface for users to enroll and obtain their certificate, this can also be done in a command line interface (CLI) which permits a batch mode for issuing several certificates to several users, we will focus on the web interface as it's the most usual way for users to obtain their certificates. EJBCA also defines methods for adding certificate extensions,for instance by editing a Java property file that specifies the certificate extensions to be used by EJBCA when issuing certificates. The following is an example of such a property file:

id1.oid=2.999
id1.classpath=org.ejbca.core.model.ca.certextensions.BasicCertificateExtension
id1.displayname=SS-Extension
id1.used=true
id1.translatable=false
id1.critical=false
id1.dynamic=true
id1.property.encoding=DERUTF8STRING

Figure 2 exemplifies the enrollment process, in the web interface, for a end user to obtain his certificate. In this example two Hospital units need to share patient records in order to manage and authorize organ donation, but also need to assure that the patient records are kept confidential and the power to share these records needs to be *"limited"* in some way. For instance consider the case where Hospital unit X has 3 Doctors responsible for organ donation and Hospital unit Y has 4, forming a group of 7, now consider that to share Electronic Patient Records (EPR) there must be an *"agreement"* between at least any 4 medical practicioners to share EPR, this scenario denotes a (4, 7) Threshold Scheme. The certificates for the various players can be obtained from the PKI, **Figure 2** shows two CA's and a suitable root certificate authority. The certificates are obtained as *p12* files since they must subsequently be imported into the clients browser or certificate store.

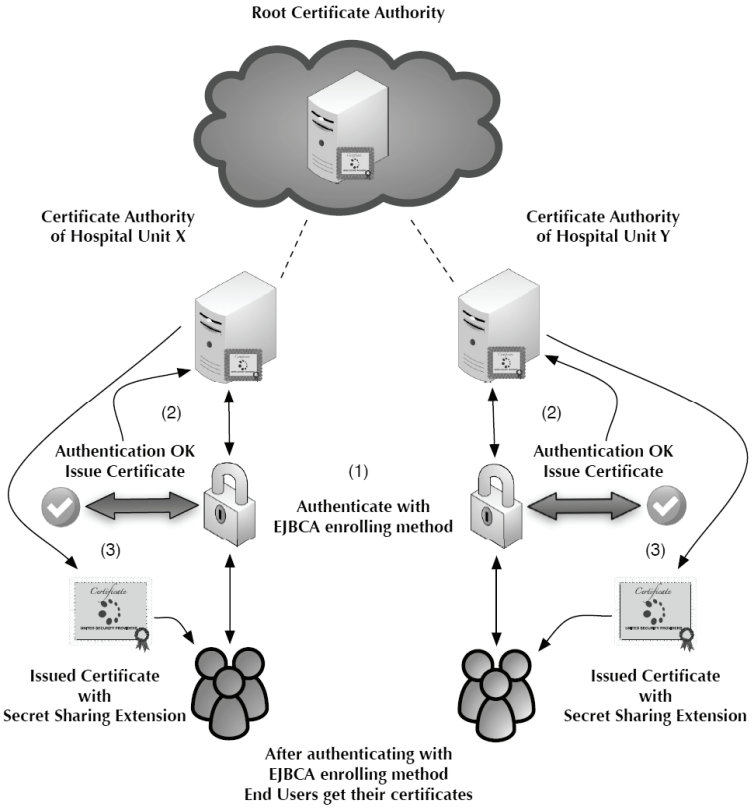
Root Certificate Authority
Certificate Authority of Hospital Unit X
Certificate Authority of Hospital Unit Y
(2)
(2)
Authentication OK Issue Certificate
Authentication OK Issue Certificate
(1)
Authenticate with EJBCA enrolling method
(3)
(3)
Issued Certificate with Secret Sharing Extension
Issued Certificate with Secret Sharing Extension
After authenticating with EJBCA enrolling method End Users get their certificates

Figure 2: Examples of end user enrollment on EJBCA to obtain certificates with secret sharing extension enabled

4. Architecture of the framework

In this section we describe the necessary requirements for the implementation of the framework, as well as some features of the PKI CA chosen as well as the communication model of the framework.

4.1 Requirements

The Basic Elements of our Framework are:

- A CA that implements a PKI according to standards such as X.509 and IETF-PKIX and supports custom certificate extensions.
- A Web Application, accessible by Web browser, where users can collaboratively form groups, share and recover secrets. This is the main application that enables end-users the "use" of Secret Sharing through their digital certificates, using the certificate extension as a security-token.

The other main requirements for the development of the framework:

- The secret and shares support integrity and owner checks;
- Access to the CA and the Web application must use strong authentication providing confidentiality and privacy;
- The framework should support implementation on a distributed orientated platform for cloud and grid computing scenarios.

It's implied that the framework must deliver a mechanism whereby the secret shares are written onto public certificates of the end users, in other words perform and update of the users public certificate. This should obviously be a flexible and easy to use system. The two possibilities are updating the users certificate on the clients local machine or reissuing the certificates from the PKI.

There are currently thousands of browser add-ons and extensions and the most user-friendly system to develop is a browser add-on/extension/ActiveX/Java Applet/Silverlight component on the client-side application that would transparently update the users certificate in the respective keystore. Unfortunately this has proved to be a difficult challenge given the wide range of browsers and operating systems and therefore the second approach was taken, namely to use the capacity of the PKI to reissue a new certificate request, this proved to be a less complex operation for the developers however it involves extra complexity for the user, as the user must repeat the certificate request, import to browser, register process. An obvious disadvantage of this approach is it invalidates the use of the certificate signature hash to sign user shares.

4.2 Implementation

Our *Secret Sharing Framework* is comprised of a Java Enterprise Edition (JEE) application deployed on Glassfish 3.1.2, a Java Standard Edition (SE) EJBCA-RA Web Service application and EJBCA, all of these applications use MySQL 5.5.x for the back-office databases. The web application is built using Java Server Faces (a Java specification for building component-based user interfaces for web applications), Primefaces (Ajax framework with JSF components), Omnifaces and the Atmosphere Framework (client and server side components for building Asynchronous Apps). iText, an open source library that allows the creation and manipulation of PDF documents is also used. EJBCA is deployed on JBOSS 5.1 (as recommended), and configured with our *secret sharing extension*. Each application has its own role in the framework which will be explained below:

- EJBCA is for PKI CA management;
- Java EE application is for authentication in the framework, group management, users choices synchronization and recover secrets;
- Java SE EJBCA-RA Web Service application is for checking group status and handling distribution of shares and management certificates status for later request by clients in EJBCA;
- EJBCA has a database in MySQL, our framework also has a database that is used by both Java EE application and Java SE EJBCA-RA Web Service.

The following **Figure 3** shows a schematic of the communication model between EJBCA-CA and the Web application using the EJBCA-RA Web service.

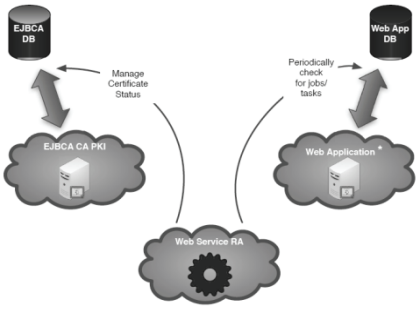

Figure 3: Web service communication model between EJBCA-CA and web application

Users need to access a *"private area"* on the *Secret Sharing Framework*, where they are able to share and recover secrets. However to access this *"private area"*, besides mutually authenticating with their certificates they must also be registered with the framework, the registration takes the form of a username which is the same as the *"Canonical Name"* in the certificate and also a password for posterior authentication. The flow of authentication in the framework is shown in Figure 4.

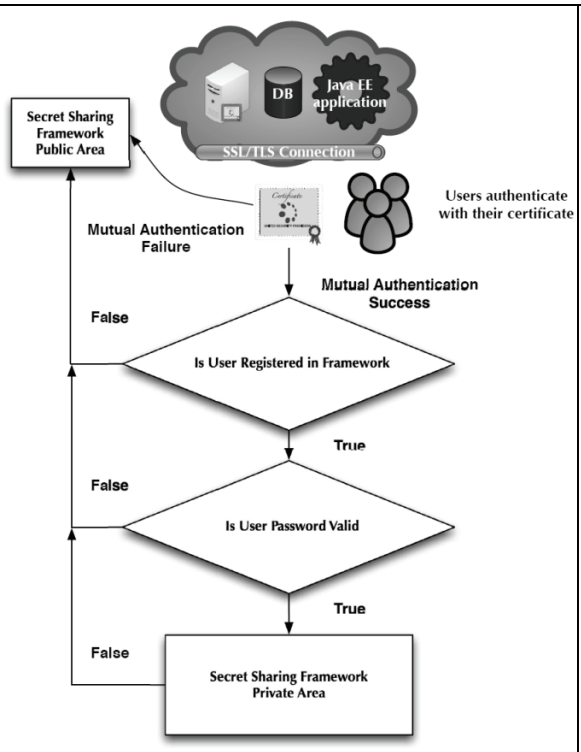

Figure 4: Flow diagram of the authentication on the framework

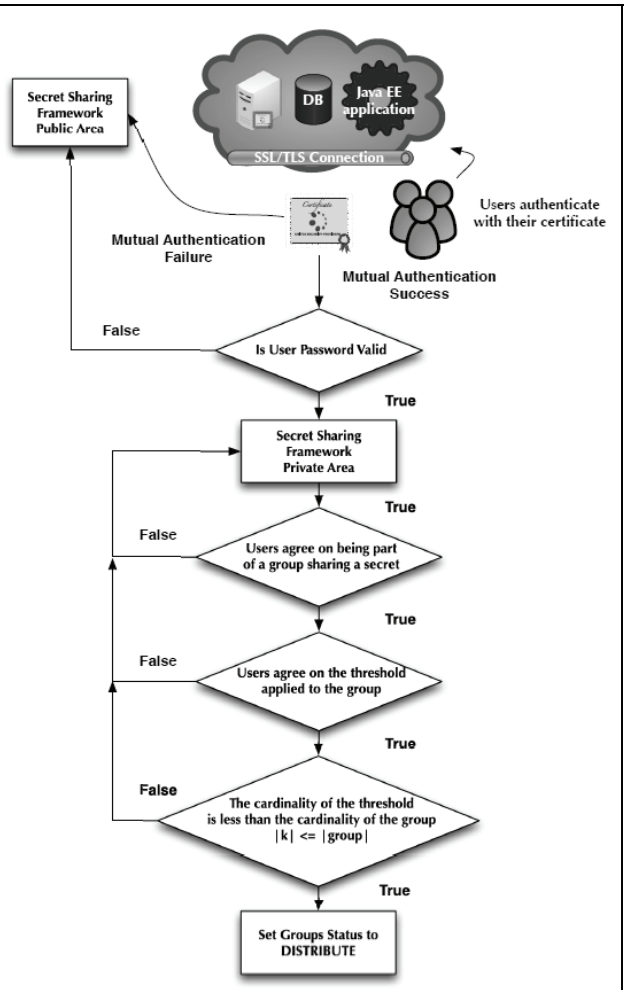

Figure 5: Flow chart for the creation of a secret distribution job/task

4.3 Secret distribution and recovery

In this section we'll show the procedures that leads to the secret distribution, as well as the algorithm implemented for secret sharing. As Shamir's SSS works on finite fields, more specifically $\mathbf{Z}_p$, we need to clarify the boundaries on this field, as all other values used by the distribution algorithm are related to these boundaries,

like the secret S that's going to be distributed and the cardinality of participants $|n|$. The most important specification is the size of *prime*, to form $\mathbf{Z}_p$, to use modular arithmetic. In our work we choose to use a 160 bit prime number p and a 156 bit secret S.

The Java EE application and Java SE EJBCA-RA Web Service share a database for group status synchronization, the groups table has a group status column with enumerated values IDLE, DISTRIBUTE, DISTRIBUTED, RECOVER, RECOVERED, BUSY, groups are initially created with the status IDLE. These values allow both applications to manage group secret distribution and recovery, the process that leads to the creation of a *secret distribution* job/task in the database that will be posteriorly caught by the RA Web service (TTP) is illustrated in Figure 5

When the trusted third party (TTP) "picks up" a new secret distribution job for a group the following algorithm 1 is processed.

Algorithm 1 - Secret Sharing Algorithm

Require: a cryptographically secure pseudo-random number generator (CSPRNG) to choose p , polynomial coefficients and S

1. TTP chooses S from $\mathbf{Z}$
2. TTP chooses $p > n \wedge p > S$ (n cardinality of users)
3. TTP selects $(k-1)$ random, independent coefficients $a_1, \cdots, a_{(k-1)}$, such that $0 \leq a_j \leq (p-1)$, defining a random polynomial over $\mathbf{Z}_p$, $f(x) = \sum_{j=0}^{(k-1)} a_j x^j$
4. **For all** $1 \leq i \leq n$ (users that belong to a group of cardinality n) **do**
    a. TTP computes $S_i = f(i) \bmod p$
    b. TTP gets *salt* from CSPRNG and computes *salted* $S_i = salt + S_i$
    c. TTP hashes SHA-256(*salted* S_i) and convert to Hexadecimal
    d. TTP updates Web Application DB with user index, hashed *salted* S_i (Hex), *salt*(Hex)
    e. **If** user certificate extension value is **null then**
        i. TTP adds values: group identifier, S_i (plaintext) and hashed *salted* S_i (Hex) to user's *certificate extension*
    f. **End if**
    g. **If** user certificate extension value **not null then**
        i. TTP concatenates values: group identifier, S_i (plaintext) and hashed *salted* S_i (Hex) to user's *certificate extension*
    h. **End if**
    i. TTP sets user client certificate status to New, in EJBCA DB, for posterior certificate request
5. **End for**
6. TTP gets *salt* from CSPRNG and computes *salted* $S = salt + S$
7. TTP hashes SHA-256(salted S) and converts *salt* and *salted* S to Hexadecimal
8. TTP gets group PDF document, in binary, from Web Application DB and encrypts PDF with AES-256 using S
9. TTP updates Web Application DB with group *modular arithmetic*, *salt*(Hex), *salted* S (Hex) and encrypted PDF(binary)

The process for secret recovery is similar. For the secret recovering process each of the participants need to log onto the framework and submit their share, each share is checked for integrity and authenticity with the hash and salt values stored in the Secret Sharing Framework database. If the total number of shares submitted by those participants is equal or greater than the previously agreed threshold, for that group, the secret can be

recovered. The shares are submitted transparently to the user being part of the certificate that the users requires in order to log onto the framework

5. Implementation and case study

The framework was implemented for the specific case where the secrets to be shared are the encryption keys of PDF documents. In this case the actual framework will store the encrypted PDF document and the document to be encrypted will be uploaded by one member of the group. The encryption key is the secret to be shared amongst the group members. The framework is available for tests at the following address http://spocs.it.ubi.pt:7070/secret-sharing/index.xhtml

Step 1: Users of the framework first need to acquire a valid certificate with our custom certificate extensions. This process is common to any certificate request process and is handled in our case by our EJBCA PKI.

Step 2: Once the certificate has been obtained users then import the certificate into their browsers (Firefox, Chrome, Safari on Mac OSX and Firefox on Windows) and can then log onto the framework. The user session is secured using SSL/TLS with Client Certificate Requirement; in this case the users will be prompted for the introduction of their private password. The Java EE server can be configured to accept certificates in various ways, certificates from a particular CA or even individual certificates. Once logon is successful the user is directed to the applications private web pages, as shown in Figure 6.

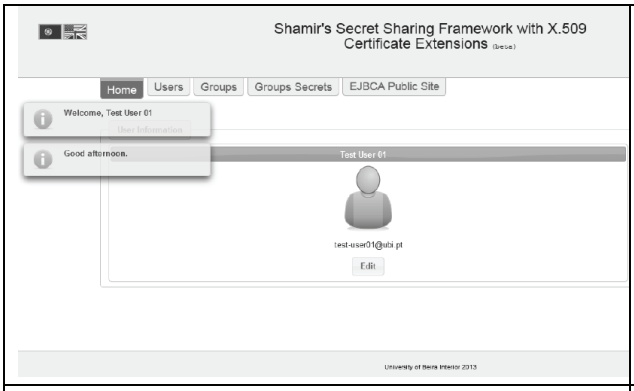

Figure 6: User logon in framework

Figure 7: Framework group creation

Step 3: Users can then update their group activity by joining a group, leaving a group or creating a new group. Figure 7 shows a user belonging to three groups in the process of creating a new group. The Framework contains a built in chat functionality to enable users to communicate with other users if necessary.

Step 4: The next step requires associating a PDF file to a group, illustrated in Figure 8 and its subsequent encryption and distribution of the secret shares as shown in Figure 9.

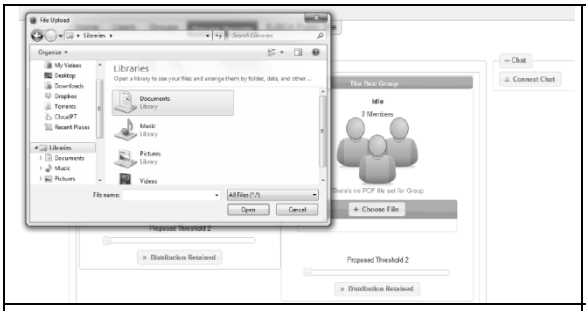

Figure 8: Get PDF file dialog

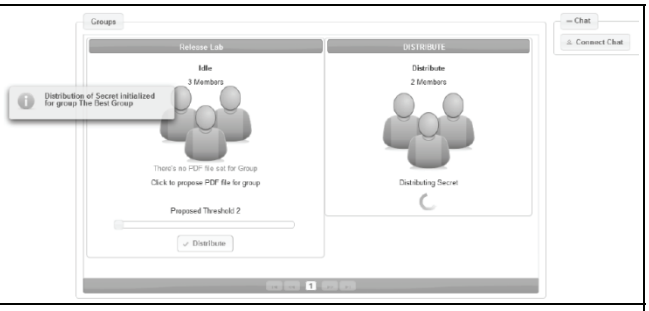

Figure 9: Secret distribution initialized

Step 5: The secret share will be then be written onto the users' public certificate. In this implementation the user gets a new public certificate from EJBCA.

Step 6: Secret Recovery involves enough group users, equal or greater to the groups threshold for the initialization of the secret recovery job/task, illustrated in Figure 10. The framework does not store the secret shares

themselves but can store metadata such as a hash of the share and owner of the particular hash, this provides mechanisms to check the integrity of the recovered share as well as preventing a share being robbed by another user and being fraudulently placed on their own certificate. After successful secret recovery the encrypted PDF file is retrieved, decrypted and shown in a separate window as shown in Figure 11

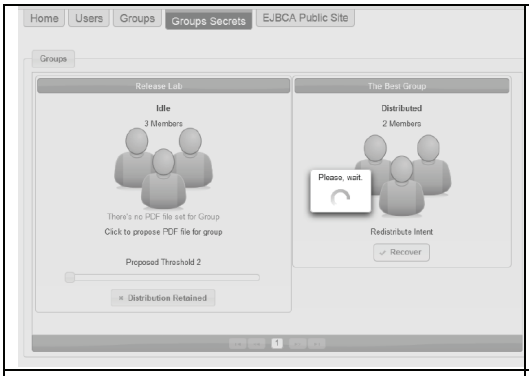

Figure 10: Recovery of shared secret initialized

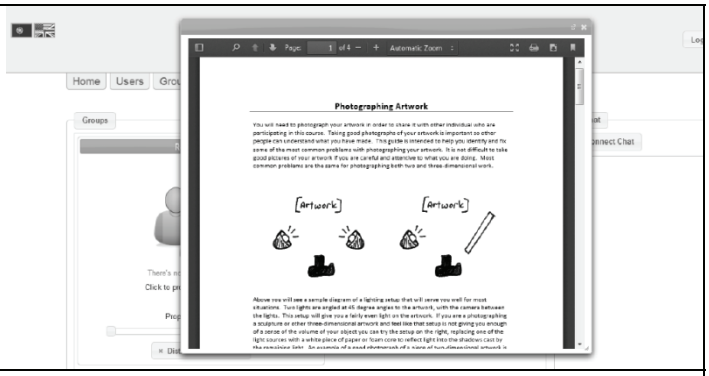

Figure 11: Shared secret (PDF) recovered

6. Conclusion

The proposed framework allows users to perform *group based operations* based on previously agreed thresholds and takes into account the privacy and confidentiality of the entities involved. The proof of concept has been tested and been shown to work correctly. Also, the framework is built using web based technologies that can be deployed in scalable and fault tolerant architectures and the user access is based on familiar browser based and digital certificate environments. In conclusion, the proposed framework enables the users to deal with collaborative group based sharing in a secure flexible environment. Future work includes extending modifying the framework to deal with hierarchy, as it is supported by a PKI, which is inherently hierarchical, this is an obvious extension and also developing client side browser components to update the certifcates with a more user friendly mechanism.

Acknowledgements

This work was partially funded by Portugal Telecom Inovação, project PRICE (Privacy, Reliability and Integrity in Cloud Environments)

References

Benaloh, J. "Secret Sharing Homomorphisms: Keeping shares of a secret." Advances in Cryptology. 1987.

Blakley, G. and Meadows, C. "Security of Ramp Schemes." Proceedings in Crypto. 1984. 242-268.

Blakley, G. "Safeguarding Cryptographic Keys." AFIPS National Computer Conference. NJ- USA: AFIPS Press, 1979. 313-317.

Brainard, J. et al. "A new two-server approach for authentication with short secrets." Proceeding of the 12th Conference on USENIX Security Symposium. Berkeley, CA - USA, 2003.

CCITT. "The International Telegraph and Telephone Consultative Comitee recomendation X.500." X500: The directory overview of concepts, models and services. 1988.

Crocker, P. and Peixinho, A. "Secret Sharing for High Security Codes on Smart-Cards." 12th European Conference on Information Warfare and Security. 2013.

EJBCA. EJBCA by PrimeKey. 2013. http://www.ejbca.org/ (accessed 2013).

Karnin, E. Greene and Hellman, M. "On Secret Sharing Schemes." IEEE Transactions on Information Theory. 1983. 35-41.

Peixinho, A. "Digital Certificates and Theshold Cryptography." Masters Thesis - University of Beira Interior. 2013.

SafeNet. 2011. http://www.safenet-inc.com/ (accessed 2013).

Shamir, A. "How to Share a Secret." Communications of the ACM, 1979: 612-613.

Yamamoto, H. "Secret Sharing System using (k,L,n) Threshold Scheme." Electronics and Communications in Japan, 1986: 46-54.

Improving Cyber-Security Awareness on Industrial Control Systems: The CockpitCI Approach

Tiago Cruz[1], Jorge Proença[1], Paulo Simões[1], Matthieu Aubigny[2], Moussa Ouedraogo[3], Antonio Graziano[4] and Lasith Yasakhetu[5]
[1]University of Coimbra, Portugal
[2]iTrust Consulting, Luxembourg
[3]Centre de Recherche Publique Henry Tudor, Luxembourg
[4]Selex ES, Italy
[5]University of Surrey, UK
tjcruz@dei.uc.pt
jdgomes@dei.uc.pt, aubigny@itrust.lu
psimoes@dei.uc.pt, aubigny@itrust.lu
aubigny@itrust.lu
moussa.ouedraogo@tudor.lu
antonio.graziano@selex-es.com
s.l.yasakethu@surrey.ac.uk

Abstract: Originally isolated by design, Critical Infrastructures (CI) based on Industrial Control Systems (ICS) – such as SCADA (Supervisory Control and Data Acquisition) systems - were born within the scope of industrial process control technologies. Having evolved from proprietary systems, ICS eventually started adopting open architectures and standards, becoming increasingly interconnected with existing corporate networking infrastructures and even the Internet. However, as these systems overcame their isolation and moved towards interconnected topologies, they also became more exposed to threats that weren't even remotely conceivable when they were first designed. Particularly, cyber-threats are one of the most significant problems that modern ICS face, as the shortcomings and vulnerabilities of the decade-old ICS technology – some of them known for a long time, but mostly downplayed due to the isolation of such systems – become serious threats that can ultimately compromise human lives. As the security needs ICT and ICS domains cannot be addressed in the same way, this calls for a domain-specific approach to cyber threat detection, designed from scratch to address its particular needs. It must consider the particular characteristics of each networking context, be it ICS or ICT, in order to provide real-time cyber-security awareness for the security teams operating in the control room – this is one of the most important contributions of the CockpitCI FP7 project (http://CockpitCI.eu), which aims at improving the resilience and dependability of CIs. In this paper we present the CockpitCI cyber-detection and analysis layer, also including a detailed description of its most relevant components in terms of role, integration and remote management. This paper will also show how the proposed solution might be effective in dealing with such cyber-threats, by presenting relevant examples.

Keywords: critical infrastructure protection, industrial control systems, SCADA systems, IDS

1. Introduction

SCADA (Supervisory Control and Data Acquisition) is the commonly designation which is used to refer a set of technologies, protocols and platforms used in Industrial Control Systems (ICS). Such systems are used in several scenarios, such as production lines automation, for controlling nuclear or thermoelectric plants, for distribution grids and many other applications.

As their scope was originally restricted to isolated environments, SCADA systems were relatively safe from external intrusion. However, as architectures evolved SCADA systems started to assimilate technologies from the Information and Communication Technologies (ICT) world, such as TCP/IP and Ethernet networking. This trend, together with the increasing adoption of open, documented protocols, exposed serious weaknesses in SCADA architectures. Moreover, the interconnection of the ICS network with organizational ICT network infrastructures, and even with the exterior (for instance, for connection with internal company systems or for remote management) brought a new wave of security problems and attacks to such an extent that the number of externally initiated attacks on ICS systems has increased significantly, especially when compared with internal attacks (Kang 2011).

This situation, together with inadequate systems lifecycle management procedures, disregarding regular updates or patching (Krutz 2006), increases the probability of a successful attack. While such procedures are

trivial matters which are part of the regular maintenance routine in the ICT world, they must be dealt in an entirely different way when it comes to ICS, mainly for two reasons: the fact that some components have to work on a continuous basis without interruptions, up to the point of working years without being reinitialized (ESCoRTS 2010) (Zhu 2011); due to the fact that any software release must be carefully tested by equipment manufacturers before being released, or even due to end-of-life support for specific devices or software frameworks.

Therefore, ICS constitute a critical and strategic asset that is being increasingly targeted by malicious attacks, with potentially catastrophic consequences. In this context, CockpitCI (CockpitCI) is focused on improving the resilience and dependability of CIs.

The idea behind the CockpitCI project is to allow the community of CI (Critical Infrastructure) owners to exchange real-time information about attacks (and tentative attacks). In this scope, each CI incorporates its own real-time Distributed Monitoring System and Perimeter Intrusion Detection System (PIDS). These systems are able to aggregate the filtered and analyzed information of potential cyber-attacks against systems used to support the operation of CIs and identify the potential unsecured area of the CIs.

This PIDS, which is the focus of this paper, performs many of the tasks traditionally associated with a Distributed Intrusion Detection System, with support for diversified and closely integrated detection and analysis techniques and tools. Each PIDS is to be deployed in the targeted area of a CI, in order to detect coordinated cyber-attacks and in order to deploy prevention strategies of isolation.

Through coordinated PIDS operations, it is possible to put in place a specific perimeter to detect potential coordinated cyber attacks on CIs for each type of detected attacks or for mixed cyber attacks. The CockpitCI PIDS is based on a state-of-the-art distributed intrusion detection architecture encompassing a set of detection agents that feed real-time and soft real-time automated correlation and anomaly detection mechanisms, orchestrated together by a management platform.

In this paper the main aspects of the proposed PIDS architecture are presented. Section 2 discusses the problem of security in ICS/SCADA. Section 3 introduces the CockpitCI architecture. Section 4 presents the proposed PIDS, with sections 5 and 6 discussing analysis components and detection capabilities, respectively. Section 7 details integration of eventing and management interfaces. Finally, section 8 presents conclusions and gives some insights into what future directions the project might take.

2. A brief overview of ICS/SCADA security issues

The development of the CockpitCI PIDS architecture was preceded by a requirements analysis phase, with the purpose of understanding the specific characteristics and differences between ICS and conventional ICT infrastructures, from a security standpoint. This study revealed several and significant differences between ICT and ICS domains that are deeply rooted in their own particular characteristics, down to the fundamental priorities that define which are the most important operational and functional properties of the system.

When it comes to their fundamental governing principles, ICS and ICT infrastructures have an inverted set of priorities, a situation that is one of the main causes of SCADA infrastructure security problems. This is partly due to the fact that, as the ICS paradigm evolved, it was not accompanied by an equal progression in terms of an industry mindset that remained unchanged almost since the inception of such technologies. As a result, and due to its critical nature, ICS operation and design practices frequently privilege availability and reliability over confidentiality and data integrity – a perspective that is quite the opposite from the ICT philosophy, which privileges confidentiality and security, followed by communications integrity and, finally, by availability (ISA-99.00.01). This contrast explains why it is frequent to find a lack of mutual understanding between the control systems teams and the security IT staff, within the same organization.

The differences between the ICT and ICS contexts also mean that there is no “one size fits all” solution when it comes to chose and implement security mechanisms. Despite this, importing solutions from the ICT world is often a necessity, which might lead to undesirable side-effects. The fundamental premises for ICT security tools and commonplace lifecycle management procedures, such as patching and updating a system, can become troublesome in an ICS, when faced with situations such as the impediment / high cost of stopping

production or even the explicit prohibition by the system's manufacturer. As an example, a SCADA system customer may be unable to install an update on an operating system unless the manufacturer certifies their software for the update.

Product lifecycle is another matter that separates ICT from ICS systems, with the former having substantially shorter lifecycles, when compared with the latter. In ICS it is frequent for mature systems to be kept in operation, sometimes far beyond their projected lifetime – the "if ain't broke, don't fix it" philosophy. This limits the possibility of implementing some security mechanisms due to the limited capabilities of existing equipment (Igure 2006).

Moreover, SCADA communication protocols, which are responsible for the interaction between field devices, such as PLC (Programmable Logic Controllers) or RTU (Remote Terminal Units) components and the stations that control and monitor them, pose security concerns. One of such examples is the Modbus protocol (Modbus 2006), originally developed by Modicon (currently part of the Schneider Electric Group) in 1979. Modbus is one of the most popular protocols for SCADA applications, thanks to its simplicity and ease of use. However, Modbus suffers from security problems: the lack of encryption or any other protection measures exposes it to different vulnerabilities (Triangle 2002) – if we take into consideration that it is not uncommon to find situations where the ICT and ICS networking contexts are blended within the corporate network, it becomes clear to which point some ICS are vulnerable. Despite this, protocols such as Modbus have a large lifespan and are still being massively deployed and used.

Simply put, when it comes to ICS, technology and platform maturity are valued as an implicit recognition of value and reliability, and even the disclosure of security issues related to them seems to have no effect in discouraging their usage or prompting the adoption of countermeasures to protect them. This has become the root cause of many ICS security issues that have been exploited with a variable degree of success, in recent times, such as the Stuxnet Trojan (O'Murchu 2011).

3. The CockpitCI project

To improve the resilience and dependability of Critical Infrastructures (CIs), the CockpitCI project proposes an architecture that is divided into several modules/components, whose interaction is illustrated in Figure 1 (CockpitCI 2013). Not only does this architecture aim to detect cyber threats using novel strategies for intrusion detection, along with devices specially conceived to monitor CI ICS/SCADA systems, but it also accounts for communication between multiple interdependent CIs, using a Secure Mediation Network (SMN) to share operational and security information.

Among the multiple components that make up the CockpitCI architecture, the Dynamic PIDS provides the core cyber-analysis and detection capabilities, being responsible for continuously assess and protect the electronic security perimeter of each CI. The automatic analysis and detection mechanisms for each PIDS are fed by several field adaptors and detection agents deployed within each CI, which constitute its "eyes", providing the basic information from which the ongoing security status of the CI is inferred. Also, the PIDS encompasses semi-autonomous reaction capabilities, being able to deploy and activate countermeasures, in line with predefined security reaction policies.

For each CI, an Online Integrated Risk Predictor (IRP) module works as a decision support system for management teams, feeding operational indicators (such as process-level data, gathered using the SCADA TLC and ELE adaptors shown on Figure 1) and cyber-security information (generated by the PIDS) into a set of modelling tools, to assess and predict propagation and threat levels for potential cyber attacks on the CI. In this scope, SCADA adaptors translate SCADA data from various components into a common data format, enabling the use of devices from different vendors and legacy SCADA HW/SW (Hardware/Software) while sharing data with the detection layer and the IRP.

The Secure Mediation Network (SMN) provides the means for exchanging security information between CIs, also enabling the use of risk prediction and other analysis mechanisms to assess threats in a global scale, accounting for CI interdependencies (as exemplified by Figure 1, which represents 2 kinds of CIs that are frequently dependent on each other: Telecommunications and Electric Power Distribution).

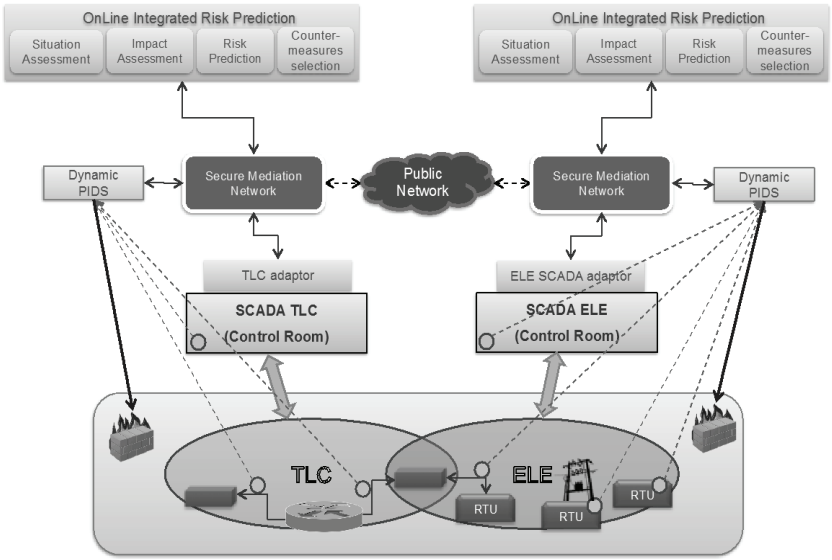

Figure 1: The CockpitCI general architecture (CockpitCI 2013)

Among these components, the CockpitCI PIDS cyber-analysis and detection architecture constitutes the main subject of this paper. The next sections will focus on the presentation and analysis of the basic components of the PIDS, explaining how they work together to continuously monitor and analyze the ICS/SCADA system components of a given ICS in order to perform threat detection. Also, the event correlation and anomaly detection mechanisms that constitute the cyber-security analysis capabilities of the PIDS will be addressed, together with the description of the agents and probes that feed such them with information about ICS/SCADA system's state and its data flows.

4. The CockpitCI cyber-detection and analysis layer within the PIDS

The CockpitCI PIDS incorporates several advanced real-time detection and analysis mechanisms, integrated to constitute a cyber analysis and detection layer for the CI, as shown on Figure 2. It is structured along the three different zones of the CI, each one with its own internal security perimeter: the Field Network, SCADA Process/Operations Network and the IT (Information Technology) Network. This distinction confers the PIDS the ability to deploy agents and security policies customized to the specific needs and characteristics of each network scope.

This architecture was designed to deal with several attack scenarios, from known threats to rogue events, such as: man-in-the-middle attacks, device impersonation, non-authorized tampering, worms, trojans, denial-of-service attacks or flooding, among others. For this purpose, the PIDS is designed in such a way that it integrates different detection strategies, distributed along different levels, namely:

- **Detection agents and field adaptors**, including agents, adaptors and extensions for existing system components, as well as specialized network probes and honeypots (Spitzner 2002) to be added to the network which are able to capture behaviour or traffic patterns (as performed by NIDS – Network IDS components) as well as host (using tools such as HIDS/Host IDS, or antivirus software) and field device monitoring.
- **A distributed multi-zone, multi-level correlation** structure that processes the information provided by the security sensors, **complemented by machine-learning capabilities**, in the form of One-Class Support Vector Machine (OCSVM) (Ma 2003) anomaly detection module, based on adaptive machine learning.
- **Aggressive usage of topology and system-specific detection mechanisms**, based on the fact that the role and behaviour of each system component in an ICS are expected to be more consistent over time than on other types of networks, analysis components are fed with knowledge provided by a number of system specific sources, such as topology databases, policy databases, and trust-based mechanisms, as well as strategically placed honeypots.

The operation of the PIDS components is orchestrated through a Security Management Platform (SMP), which is responsible for managing all the involved components of the solution (see Figure 2). It includes the

mechanisms for managing the security and components of the infrastructure. The SMP is responsible for the maintenance and management of monitoring probes such as IDS and the analysis components, also including monitoring of in-place security and vulnerabilities within the network as well as the maintenance of the latter. Therefore, the SMP has a dual role, dealing with both security audit and maintenance mechanisms.

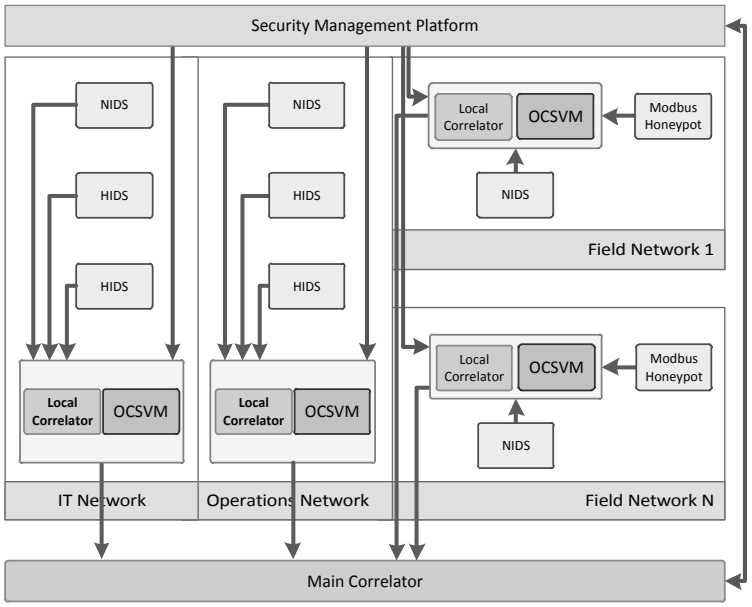

Figure 2: The CockpitCI cyber-detection and analysis layer (red flows=management, green=eventing)

The SMP performs the configuration of detection agents on the field, allowing to set-up their detection thresholds and other relevant parameters. This detection threshold depends on both the risk level of the overall infrastructure (the level of detection shall be higher if the probability of an attacks is higher) and on the specific detection needs (e.g. in case of abnormal event detection on specific system, the SMP shall be able to verify all similar components in the CIs to check their security level).

Due to the demanding availability requisites and little tolerance to delays, the detection architecture is to be implemented using a network that is separate from the SCADA system network (eventually it can use the same physical network, using VLAN (Virtual Local Area Network) or other types of overlay techniques for traffic separation), in order to guarantee that it does not interfere with the normal operation of the control network.

5. Analysis layer

The analysis components of the PIDS provide a way to extract information from the data collected by the agent layer or directly from network traces. These components are arranged in a two-level architecture with local instances fine-tuned for each network scope.

5.1 Local and global correlators

Local correlators perform the first step of correlation, filtering and reducing the number and noise of the alarms generated by the detection layer, while providing a mechanism for security event generation that is able to filter, process and relate events within a network segment (e.g. alarms generated by two or more detection agents, multiple events from the same source).

Local correlators receive the events from local detection agents (e.g. HIDS) on their network scope and process them accordingly with a set of rules, forwarding significant results to a global correlation engine. This approach provides context separation, at the same time allowing for better efficiency and scalability for real time event processing. After local correlation, events are sent to the global correlators and from the latter to the SMN, using the Intrusion Detection Message Exchange format (Debar 2007). IDMEF defines an experimental standard for exchanging intrusion detection related events. As a standard, it can be used as a vendor or product independent enabling intercommunication between different agents such as NIDS or Honeypots.

As illustrated by Figure 2, local correlators receive events from the different agents such as NIDS, HIDS, Honeypots, among others. These agents are distinct according to network zone in which the local correlator is located. Despite of the range of different agents, the local correlator should use the same interface for all of them, as messages are received through an Event Bus (discussed on Section 7). This interface will allow subscribing to the events published by the agents. Local correlators also have an agent adaptor interface that allows for management, via the SMP.

Regarding the event interfaces for the main correlator we have different types: one to receive events from the local correlators and another one to send events to the SMP, both using an Event Bus. As local correlators have already previously processed received events, the main correlator can focus in Multi-Step, Attack Focus Recognition correlation, as well as Alert Prioritization. A management adaptor provides the interface for the SMP to configure the correlator (see Figure 3).

The correlators are implemented using the Esper (Esper) Complex Event Processing (CEP) tool. This was due to the fact that Esper is a multi-platform, flexible and mature tool, in development since 2006. Also, performance tests have shown Esper to exhibit a good balance between memory usage, CPU usage and execution time, when processing hundreds of thousands of events.

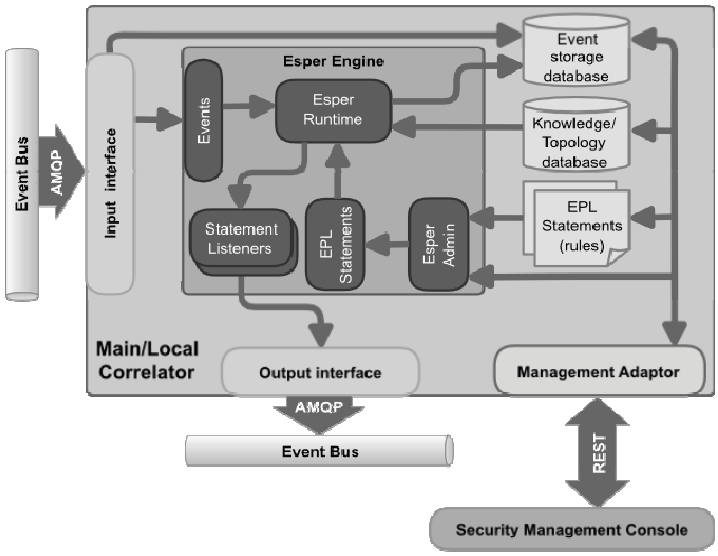

Figure 3: verview of the correlator architecture

Esper can natively accept events represented in XML, among others, which is useful as IDMEF, used by the PIDS, is an XML-based format. If a XML schema document (XSD file) is provided Esper can read the schema and properly present event type metadata and validate statements that use the event type and its properties. To access the elements of the event the correlator uses XPath expressions. If a schema for the XML is provided, the XPath expression needed to reference the attribute can be inferred automatically. Otherwise, expressions can be manually configured.

An overview of the architecture of a PIDS correlator, based on Esper, is pictured in Figure 3. The events received from the input adaptor are sent to the Esper Runtime (EPRuntime); this provides the interface to the event stream processing runtime services. The statements are registered in the EPRuntime and represent the event stream queries and/or event pattern. Each statement can have one or more Statement listeners bound to them. When the condition of a query is verified Esper can trigger the listener(s) bound to the rule, insert the result of the statement into another stream (that already exists or is created at that time) or do both options. If a rule generates a new event, that need to be sent by the correlator to the Event Bus, the listener will interface with the output adaptor to send it. Output events are generated by rules making use of input events, cached events, the internal state and information from external sources.

Esper statements are added to the EPRuntime through the Esper Administrator (EPAdministrator) module. This is an administrative interface to the event stream-processing engine. For security auditing purposes the correlator will log all events and traces of the actions performed to persistent storage. The events will be logged as they are received in the correlator and the EPRuntime shall also log the actions executed by the

correlator. Correlation can make use of information taken from external sources. These sources can provide additional information related, among others, to the definition of the network topology and other detailed system information. These external sources (knowledge/topology databases) can be queried directly from an EPL statement. New rules can be added to the correlation engine dynamically, without restarting the engine.
Using the same correlator tool for the two levels of correlation provides uniformity, since the same language is used to express the correlation operations, and allows easier integration with the Event Bus, as the same interfaces can be used for the two levels. Using the same rule description language for both correlators simplifies the task of rule management by operators and security experts. Additionally, some correlation rules can be used in both correlators without the need to be converted.

5.2 One-Class support vector machines (OCSVM)

OCSVM (One-Class Support Vector Machine) are a natural extension of the support vector algorithm to the case of unlabelled data, especially for detection of outliers. However, unlike SVM or any another classification algorithm, OCSVM does not need any labelled data for training or any information about the kind of anomaly is expecting for the detection process. OCSVM principles have shown great potential in the area of anomaly detection (Ma 2003, Li 2003, Schölkopf 2001). Moreover, OCSVM is capable of handling multiple attributed data (Hsu 2003, Wang 2004), which is well suited for SCADA systems.

The advantages of the OCVSM component are manifold: since OCSVM does not require any signatures of data to build the detection model it is well suited for anomaly-based intrusion detection in SCADA environment; since the detection mechanism does not require any prior information of the expected attack types, OCSVM is capable of detection both known and unknown (novel) attacks, besides being robust to noise in training sets. Also, algorithm behaviour can be controlled and fined-tuned by the user to regulate the percentage of anomalies expected (thresholds, as defined via SMP via the OCSVM management adaptor).

OCSVM operation consists of 2-steps, namely: training and testing. During the training stage OCSVM builds a model from training on normal (i.e., obtained from a system operating under normal conditions, without any attack in progress) data and then classifies the new data as either normal or attack based on its geometrical deviation from the training data in the testing stage. Since the OCSVM detection approach is robust to noise samples, the training data set can include some noise samples (i.e. data which does not correspond to the normal behaviour). An OCSVM component is deployed in IT, Operation and Field network zone(s), therefore requiring different training sets.

Once the training phase is complete, the OCSVM module is capable of detecting possible intrusions (abnormal behaviour) to the SCADA system, based on real-time capture of network traffic traces. The detection module will classify each event whether it is a normal event or a possible intrusion. This information will then be encoded in an IDMEF message and sent to the main correlator, using an adaptor for the Event Bus, in order to react accordingly to the detected intrusions.

6. Detection agents

The detection agents are the lowest level of the detection layer. Their purpose is to gather information from the system. As the format of information provided depends on the type of detection agents used (type of probe), adaptors allow the acquisition of data from the system in a recognised format. Detection agents and adaptors are essential to feed the local correlators of the detection layer with input data regarding suspicious activity. The PIDS encompasses several kinds of probes and detection agents, among which the most relevant are next described.

6.1 Threat detection agents

Network IDS: the perimeter for each network scope is monitored using NIDS components for each one: IT Network NIDS, Operations Network NIDS, and Field Network NIDS. These have interfaces to report the security events to the zone correlator within their network scope. In the PIDS, Snort (Snort) is used for this purpose, albeit other NIDS could be used.

Host IDS: the Host IDS is deployed in the hosts/servers of the system. It is capable of reporting anomalous behaviour in the machine where it is deployed. In the CockpitCI PIDS, OSSEC (OSSEC) is used for this purpose, but other HIDS could be used.

Honeypots: acting as decoys and being capable of detecting attackers probing the network, honeypots provide another source of data for correlation. There are three types of honeypots in the detection layer: IT Network, Operations Network and Field Network honeypots (Simoes 2013).

Exec Checker (linux hosts): capable of detecting malicious network frames by sniffing the traffic, the Exec Checker (in active or passive mode) captures the different parts of an executable in the network traffic to recreate the file and send it to an analysis tool.

6.2 Vulnerability detection agents

Output Traffic Controls (linux hosts): capable of detecting Remote Access Trojans, this specific tool regularly scans system components to check if a remote access toolbox has been installed on components to facilitate external attacks.

Vulnerability Checker (windows hosts): this tool provides a regular control of system vulnerability to check if the monitored systems are vulnerable or not according to an updated database. This tool can be customized for IT or SCADA host profiles.

Configuration Checker (linux/windows hosts): this tool provides a regular control of system configurations to check for unauthorized modification.

6.3 Security event detection agents

Behaviour checker (linux/windows hosts): capable of detecting attacks/threats by analyzing low-level hardware/software behaviour, this specific family of detection agents retrieves hardware/software information such as temperature and CPU (Central Processing Unit) activity in order to avoid accidental or malicious outage.

Security events generated by detection agents are encoded using the IDMEF format. All detection agents have a separate channel (another interface or secure channel) for management purposes, enabling the security staff to adjust the configurations with the scenario requirements, via the SMP. The detection agents send their messages by means of an Event Bus described in Section 7, which also details the management interfaces for the agent adaptors. These interfaces (eventing and management) were designed to ease integration of several types of detection capabilities (such as antivirus, for instance) providing wrapper components for event generation and the management API.

7. Interfaces and integration

This chapter describes the transport mechanisms and interfaces for event data flowing between the several existing components of the PIDS, also addressing their management interfaces.

7.1 The event bus

The Event Bus is the component responsible to manage the communication of the events between the different elements of the PIDS, whose architecture is detailed in Figure 4. Events generated by the different agents within each zone are sent to an Event Bus broker. The broker is then responsible to route this events to a queue from which the local correlator can consume them. After processing and correlating, the events each local correlator sends the events to another broker that feeds the main correlator. The events produced by the main correlator are sent to the main broker that routes them to a queue where they can be sent to the SMP.

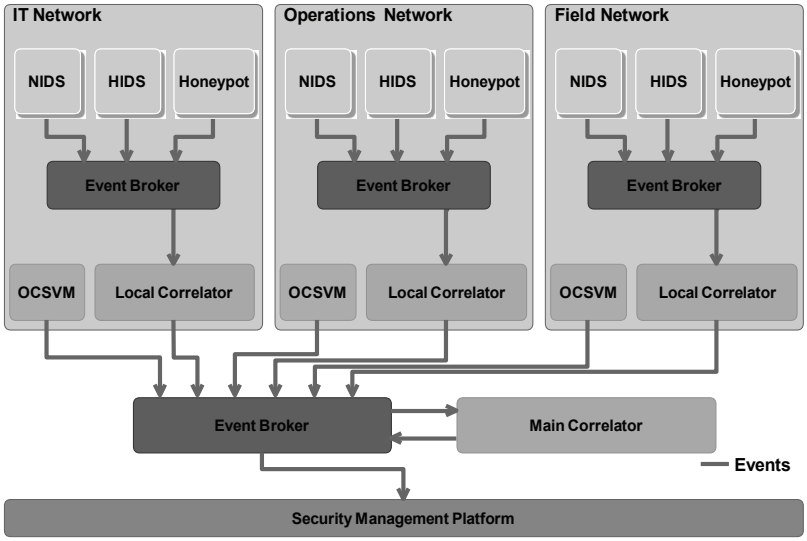

Figure 4: Event bus architecture

The Event Bus uses a Message Oriented Middleware (MOM) (Banavar 1999) to provide efficient event communication among the (sometimes, heterogeneous) components that comprise the PIDS. Several MOM implementations depend on a Message Queue (MQ) system to allow asynchronous message delivery, by providing a temporary storage, on memory or disk, for the messages. Messaging applications communicate with each other through a messaging system, acting either as a message producers (senders) or consumers (receivers). Producers and consumers are loosely coupled, being connected through virtual channels called publish-and-subscribe (one-to-many) channels or point-to-point (one-to-one) channels (Chappell 2004).

For the integration of eventing interfaces, the CockpitCI PIDS adopted an Event Bus based on the Advanced Message Queuing Protocol (AMQP) (OASIS 2011), a wire-level, open standard application layer protocol for MOM that defines a neutral (IDMEF-compatible) encoding scheme of byte sequences to pass over the network. An AMQP messaging system comprises three main components: publisher(s) (which assemble messages and send them to a message queue) consumer(s) (which receive messages from a message queue) and broker(s)/server(s) (responsible for receiving messages from publishers and route them to the right consumers).

The AMQP-based MOM brings a set of important features to the PIDS architecture, namely:

- **Security:** it supports authenticated and/or encrypted transport, using Transport Layer Security (TLS) or Simple Authentication and Security Layer (SASL), to protect events from tampering and/or eavesdropping.
- **Message reliability:** it can guarantee message ordering using a queuing broker, ensuring that messages are delivered to the receiver in the same order in which the sender sent them, with support for disconnection (messages may be held in a queue for deferred delivery).
- **Resiliency:** message delivery semantics provide a range of delivery options, with special emphasis to the *exactly-once* and *at-least-once* modes. These delivery modes, guarantee the message to arrive to the intended destination no matter what. The messaging provider will retry the delivery of a message upon a delivery failure.
- **Scalability and High Availability:** it provides scalability for the communication system thanks to the publisher-subscriber model. The agents can send events, publishing them to a queue/exchange in the broker, which is subscribed by a correlator to receive the messages. This allows adding additional consumers with ease, for failover or to distribute the correlation load across more than one instance. Also, a group of brokers can be clustered together for high availability and/or scalability/load-balancing.

Moreover, the protocol is vendor-neutral and platform-agnostic. There are several open source implementations for many different programming languages.

7.2 Management interfaces

For each managed entity that does not provide a suitable management interface, a component management adaptor/coupling architecture provides an uniform API and Data Model for each component that does not expose its own native management interface.

The Management Adaptor also embeds an API/Data Model module that is responsible for maintaining its data model (state and semantics) properties and also to provide the web service API interface to manipulate them. Accordingly with the mapping rules from the Abstraction Class, attributes exposed by the API layer might have several properties, defining and describing their access mode (read, write), or data types. The API makes use of REST (REpresentative State Transfer) (Fielding 2000) web services, with security being provided with the help of HTTPS with other authentication mechanisms such as client certificates or signed requests.

The data model structure for management adaptors is standardized, being inspired on hierarchical models usually found on management protocols such as SNMP (Case 2002), being arranged as a tree. Asynchronous events are also supported though inclusion of eventing properties, enabling a specific attribute to generate notifications when its state changes.

8. Conclusion

This paper presents the architecture of the PIDS within the CockpitCI architecture. This architecture was designed to address the special cyber-security needs of CIs, such as ICS/SCADA systems, being based on a distributed approach that attempts to bring the most effective detection mechanisms and tools together with correlation and anomaly detection analysis techniques, in order to create a solution that starts with the state-of-the-art in CI security as its baseline.

A strong point of this architecture lies in its capability for assimilation of a diverse range of detection tools in a coherent framework with homogeneous coordination and orchestration. Using distributed two-level correlation capabilities the PIDS is able to get a micro and macro-perspective on the ongoing status of the monitored CI, while being capable of dealing with unknown threats, thanks to the incorporation of machine-learning anomaly detection features. Future work will address improved integration with the SMN, while expanding on functionality and diversity of detection components.

Acknowledgements

The authors would like to thank the support of the CockpitCI (FP7-SEC-2011-1 Project 285647) and iCIS (Intelligent Computing in the Internet of Services – CENTRO-07-0224-FEDER-002003) projects.

References

OASIS, Advanced Message Queuing Protocol (AMQP), version 1.0, available at: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=amqp, July 2011.

Banavar, G., Chandra, T., Strom, R. and Sturma, D. (1999) A Case for Message Oriented Middleware, IBM T. J. Watson Research Center, Hawthorne, New York.

J. Case et al. (2002) Introduction and Applicability Statements for Internet Standard Management Framework, IETF RFC 3410, December 2002.

D. Chappell (2004) Enterprise Service Bus, O'Reilly Media, 2004

CockpitCI, CockpitCI FP7-SEC-2011-1 Project 285647, available at: http://CockpitCI.eu.

CockpitCI (2013) CockpitCI FP7 Deliverable D3.1, Requir. and Reference Arch. of the Detection Layer.

H. Debar, D. Curry, B. Feinstein (2007) Rfc 4765: The intrusion detection message exchange format (IDMEF), March 2007, http://www.ietf.org/rfc/rfc4765.txt.

ESCoRTS (2010), TAXONOMY of SECURITY SOLUTIONS for the SCADA Sector, Deliverable 2.2,

Esper Complex Event Processing, EsperTech, available at: http://www.espertech.com/products/esper.php.

Fielding, R.T. (2000) Architectural Styles and the Design of Network-Based Software Architectures, Ph.D. Dissertation, University of California, Irvine.

Hsu, C., Chang, C. and Lin, C. (2003) A practical guide to support vector classification, Technical report, Dept. of Computer Science and Information Engineering, National Taiwan University, Taipei.

Igure, V.M.; Laughter, S.A. and Williams R.D. (2006) Security issues in SCADA networks, Computers; Security, Volume 25, Issue 7, Pages 498-506, 2006.

ISA-99.00.01 (2007) Security for Industrial Automation and Control Systems - Part 1: Terminology, Concepts, and Models, American National Standard.

Kang, D. et al., (2011) Proposal strategies of key management for data encryption in SCADA network of electric power systems, Int. Journal of Electrical Power & Energy Sys., Vol. 33, Iss. 9, Nov. 2011.

Krutz, R. L. (2006) Securing Scada Systems, USA: Wiley Publishing, Inc., 2006.

K. Li, H. Huang, S. Tian and W. Xu (2003) Improving one-class SVM for anomaly detection, Proceedings of the Second Int. Conference on Machine Learning and Cybernetics, Xi'an, 2003.

J. Ma and S. Perkins (2003) Time-series novelty detection using one-class support vector machines, Proceedings of the International Joint Conference on Neural Networks, July, 2003, pp. 1741-1745.

Modbus-IDA (2006) Modbus Application Protocol Specification V1.1b.

L. O'Murchu, N. Falliere (2011) W32.Stuxnet dossier, Symantec White Paper, February 2011.

OSSEC, Open Source SECurity, Trend Micro, available at: http://www.ossec.net.

B. Schölkopf, J. Platt, J. Shawe-Taylor, A.J. Smola, and R. Williamson (2001) Estimating the support of a high-dimensional distribution, Neural computation, Vol. 13, No. 7, pp. 1443-1472, 2001.

P. Simões, T. Cruz et al. (2013) On the use of Honeypots for Detecting Cyber Attacks on Industrial Control Networks, In proc of 12th European Conf. on Information Warfare and Security (ECIW 2013).

Snort IDS, Sourcefire, available at: http://www.snort.org.

Spitzner, L. (2002) Honeypots: Tracking Hackers, Addison-Wesley Professional.

Triangle MicroWorks, Inc (2002) DNP3 Overview, Raleigh, North Carolina, http://www.trianglemicroworks.com/documents/DNP3_Overview.pdf.

Y. Wang, J. Wong, and A. Miner (2004) Anomaly intrusion detection using one class SVM, presented at 5th Annual IEEE Information Assurance Workshop, West Point, New York, 2004.

Zhu, B et al. (2011) A taxonomy of Cyber Attacks on SCADA Systems, Proc. of the 2011 Int. Conf. on Internet of Things and 4th Int. Conf. on Cyber, Physical and Social Computing (ITHINGSCPSCOM'11).

Putting Counterintelligence in Cyber Counterintelligence: Back to the Future

Petrus Duvenage and Sebastian von Solms
Centre for Cyber Security, University of Johannesburg, South Africa
duvenage@live.co.za
basievs@uj.ac.za

Abstract: Businesses and governments alike are grappling with configuring their cyber security postures effectively in a manner to account for rapid changes in the cyber threatscape. Defensively security policies and measures (inclusive of software and hardware technologies) on their own are wholly inadequate in protecting against proliferating threats. An effective approach for securing and advancing cyber interests will have to combine more proactive defences with intelligence on, and the engagement of, adversaries. Offensive and defensive measures, in turn, should be integrated with an institution's strategy and objectives. Appropriately conceptualised, Cyber Counterintelligence (CCI) could meet these requirements and offer a practicable approach for governments, businesses and other sizable entities. There is a precondition. To be effective, CCI should be an integral part of multi-disciplinary Counterintelligence (CI) – conceptually and in practice. However, at least in as far as consulted academic literature is concerned, such conceptualisation is lacking. Disconcertingly, the theoretical discourse about CCI is gaining momentum without a categorical explication of CI. Such theory construction is flawed and, within CI, the price for bad theory is paid in costly failures. Therefore, this paper conceptualises CCI as part of CI. To this end a cursory primer on CI is provided. Building on this primer, the paper proceeds with advancing: (i) A definition of CCI. (ii) A three-tiered postulation for conceptually integrating CCI with multi-disciplinary CI, Intelligence and Strategy. (iii) A taxonomy of CCI tools, methods and means. (iv) A matrix that has the dual purpose of (a) categorising CCI tools, methods and means; and (b) plotting a CCI posture in accordance with the nature and the needs of a specific institution.

Keywords: counterintelligence, cyber security, cyber counterintelligence

1. Introduction

What was seen as a paradigmatic shift in thinking at the turn of this decade is now commonly accepted – conventional cyber security we have been relying on is deteriorating on all fronts (Lües 2012). As a result, cyber space is now probably more insecure than it has ever been (Bodmer et al 2012). It is also likely to be the most secure than it is going to be for the foreseeable future. It is, simply put, going to get much worse. According the World Economic Forum (WEF 2014): "[t]he world may be only one disruptive technology away from attackers gaining a runaway advantage." The past two years has indeed witness an unprecedented escalation in the sophistication and detrimental impact of cyber criminals, hacktivists and other such role-players. Simultaneously, the pervasiveness of nation states' cyber surveillance by the intelligence apparatuses of not only the United States (U.S.) and United Kingdom (UK), but also the People's Republic of China (PRC) and Russia continue to make headlines.

With the awareness of conventional cyber security's faltering, both state and non-state actors have been intensifying the quest for ways to more effectively protect and advance their cyber interests and, in the case of service providers and vendors, those of their clients. As could be expected, solutions offered in the market place vary considerably. Buzzwords and marketing slogans currently gaining favour include counter exploitation, threat intelligence, offensive measures, hacking back, threatscape and intelligence software analytics (IBM 2013, Helton 2013). Common to most of these solutions advocated, is recognition of the imperative of intelligence on threats actors and the need to engage threats pro-actively/offensively. There is a sense of taking the fight to adversaries. The use of such notions and phrases would have been encouraging if it was indicative of a move to proactive, integrated and balanced approaches (Helton 2013). As it currently stands the use of these terms is, however, disturbing for three inter-related reasons. Firstly, the terms are used vaguely and without the proper context from the statutory intelligence practice from which they are often derived. In marketing jargon 'intelligence' is used interchangeably with 'data and information.' Consequently, solutions offered under the rubrics of 'intelligence' may not solve the problems for which they purport to be the fix. In a similar vein the opting for quick-fix offensive actions, not dovetailed with an appropriately configured defensive posture, is inviting disaster. Equally disconcerting, solutions and terms are sometimes thrown around without qualifying that some aspects of cyber defence and offense are the exclusive prerogative of the state apparatus in most countries. These functions ought not to be 'out-sourced'

to other entities. Secondly, solutions being offered are essentially technical and tactical in nature. As important as they are, technical and tactical measures on their own, were mentioned earlier, as being insufficient to confront the sophisticated threat actors about whom we are most concerned. Social engineering, to cite one example, "played a part in nearly every major hack or breach in 2013 yet it still stays in the background when we consider security controls. This is something that needs to change as we move forward." (ISC^2 2013). Thirdly, these solutions are presented as neat 'add-ons' or 'plugins' to be used as a layer additional to existing cyber security measures. 'Add-ons' seldom have, and certainly will, not in future, offer adequate protection against advanced adversaries. To counter adversaries, cyber security has to be part of organisational DNA and not mere feel-good plasters offering little real protection.

There is a way in which such synergy can be achieved and the tables turned on cyber adversaries. This paper posits cyber counterintelligence (CCI) as a practicable approach to effectively securing and advancing cyber interests. From this perspective, malicious cyber actions are not all bad news. The good news is that we can exploit malicious cyber actions to our advantage and adversaries' detriment. There is, however, a precondition and there are no half measures. To be effective CCI needs to be properly conceptualised and implemented. If not, it is your highway to hell that could end in self-destruction. For a substantial part, this conceptualisation entails the application of time-tested counterintelligence (CI) notions to the cyber sphere. It is a case of going back to time-tested CI fundamentals in order to enable our prosperity in the cyber sphere today and in the future. It is a case, as suggested by the paper's title, of going back in order to successfully move to a more secure cyber future.

This paper's primary aim is to provide a conceptual baseline that could aid in stimulating the academic discourse on CCI. Consequently, it ensues with a cursory overview of the status of CCI in the public and academic discourse on cyber security. An academic self-awareness of CCI under-theorised status is a first step in addressing this near void. The paper proceeds with advancing CI and CCI constructs hopefully useful to this discourse. Rather than aiming to impart radically new concepts, the emphasis is on presenting existing knowledge in a manner conducive to further academic debate. Such conceptual 'building blocks' include a definition and delineation of the CCI as a CI sub-discipline, a taxonomy for CCI methods and means as well as a CCI matrix for configuring an offensive-defensive posture. It concludes with some views on CCI's future and by suggesting areas for further academic enquiry.

As noted above, the following section reflects on CCI under-theorised status, since such awareness is an important first step in make progress with the academic discourse.

2. Cyber counterintelligence's under-theorised status

In one form or the other CCI has been practised as part of the statutory CI functions in various intelligence communities for well over two decades. CCI has also been offered as service by a few niche companies for some time. Until very recently, however, CCI has not really gained traction outside the statutory security structures and the relatively small batch of clients the niche companies served. Despite the key it holds to secure cyber interests for state and non-state actors, CCI entered the second decade of the 21st century underappreciated and underexplored in policies and other literature in the public domain. The overwhelming majority of governments' cyber security policies do not make any references to counterintelligence. In the few instances that the concept is cited, counterintelligence is hardly at the centre.

Nevertheless, CCI has fast been gaining traction during the last two years. The 2013 proceedings of the 12th European on Cyber Warfare and Security (ECCWS), for example, comprise of 44 papers and is a voluminous 406 pages in length (Kuusisto & Kurkinen, 2013). Not one of these papers makes mention of CCI. In fact, only one sentence in the whole of the proceedings makes cursory reference to the concept counterintelligence (Kuusisto & Kurkinen, 2013: 244). One year on and ECCWS 2014 features a dedicated mini-track to Cyber Intelligence/ Counterintelligence. While this certainly reflects an increased awareness of CCI, contributions are scarce. This is one of only two papers at ECCWS 2014 with CCI as a primary focus.

While a few commendable books have been published on the subject, these are minuscule in comparison with the proliferating material on cyber security generally. However, the shift in emphasis towards CCI is also apparent here. An outstanding work by Bodmer *et al* was first published in 2012 with the title *Reverse*

Deception – Organized Cyber Threat Counter- Exploitation (Bodmer et al 2012). The updated edition, due for release in 2014, is more aptly entitled *Hacking Back: Offensive Cyber Counterintelligence* (Bodmer et al 2014).

Although CCI is set to gain in prominence, the participation in and agenda of this discourse will inevitably be influenced by the relative obscurity of CI in general. CCI will be demonstrated in further sections as a sub-set of the broader, multi-faceted CI discipline. It thus follows that contributions to CCI would need to be preceded by some grounding in CI. CI, however, and herein lies the glitch, is in itself academically obscure and underappreciated. This obscurity is as old as its inception as a formalised discipline in the run-up to the Second World War. One reason for CI's obscurity in the academic debate is its complexity. The following statement by a U.S. CI veteran during the Cold War has lost none of its relevance: "It is not easy, nor can one feel confident, to re-enter this world where, it has been said, the tortuous logic of counterintelligence prevails ... Unfortunately, there seems to be no easy way to explain counterintelligence ... Because effective counterintelligence is a combination of so many aspects." (Miller, 1980) Even among policy makers, scholars and "national security practitioners" in foremost intelligence communities such as those of the U.S., "the role of counterintelligence remains little known or understood" up to this day (Van Cleave 2007).

Since the role, functions and importance of CI is opaque within statutory intelligence circles, the reluctance of 'techies' to apply this concept to the cyber sphere is understandable. In a similar vein, those skilled and experienced in more conventional counterintelligence do not necessarily have a sound working knowledge of technical cyber. If we are not clear on a conceptual level, we can hardly make progress in the academic discourse, thereby eventually affecting our ability to implement sound solutions. In the conceptual difficulties of CI also lies the opportunity. If we can understand and explain CI, we can explain CCI and unlock the latter's potential as a force multiplier.

This section illustrated the need for contributions to the budding field CCI to be clearly rooted in CI. In line with this contention, the next section provides a conceptual primer of CI.

3. A primer on counterintelligence

Counterintelligence has been practised for millennia. Some enduring principles were penned in 500 B.C. by the much-quoted Sun Tzu in a specific chapter in his *The Art of War* devoted to the use of spies and counter-spies (Giles 2002). The term in its contemporary connotations entered the English lexicon in the mid-1930's (Dictionary.com 2013). For some, counterintelligence is all about spies chasing spies. For others, it invokes mundane security measures. Counterintelligence is both of these aspects and so much more.

3.1 Delineating counterintelligence

Counterintelligence can be defined as the collective of measures undertaken to identify, deter, exploit, degrade, neutralise and protect against adversarial intelligence activities deemed as detrimental or potentially detrimental to the own interests (Duvenage 2010). The term 'counterintelligence' is thus an abbreviated form for the *counter*ing of hostile *intelligence* activities. Adversaries engaging in hostile intelligence actions include nation states, corporate entities, criminals, activists, individuals and any combination of these.

Adversarial intelligence activities comprise espionage, deception (disinformation), influencing and some forms of covert action. Of these different intelligence activities, espionage is the most central. Espionage to obtain protected information in order to gain a competitive advantage can be an end in itself; or such information can be used to further other malicious ends such as data manipulation, disinformation and disruption. Sophisticated adversaries execute their intelligence actions through the exploitation of humans (HUMINT) and technical means (TECHINT). The latter, in turn, comprise Signal Intelligence (SIGINT), Imagery Intelligence (IMINT), Measurement and Signature Intelligence (MASINT) and Cyber Intelligence (CYBINT). These conduits and their relation to adversarial intelligence ends can graphically be depicted as follows in Figure 1.

Several revelations by the whistle-blower Edward Snowden provide a practical illustration of the above. During September 2013, for example, *The Guardian* newspaper revealed that the British and U.S. Agencies run HUMINT operations to "help secure an insider advantage" (Ball et al 2013). To this end the British Government Communications Headquarters (GCHQ) established a HUMINT Operations Team (HOT) "responsible for identifying, recruiting and running covert agents in the global telecommunications industry." (Ball et al 2013). These operations enabled the Agency to "tackle some of its most challenging targets" – specifically in the

encryption sphere (Ball et al 2013). In this instance, success in the field of CYBINT thus depended on the effectiveness of HUMINT operations. The reverse is of course also true. Given the high and growing digital dependence, CYBINT is often a critical enabler in the HUMINT sphere. To be effective CI needs to counter all types of adversarial intelligence activities and, in the case of high-end adversaries, it has to do so in more than one or all of the conduits.

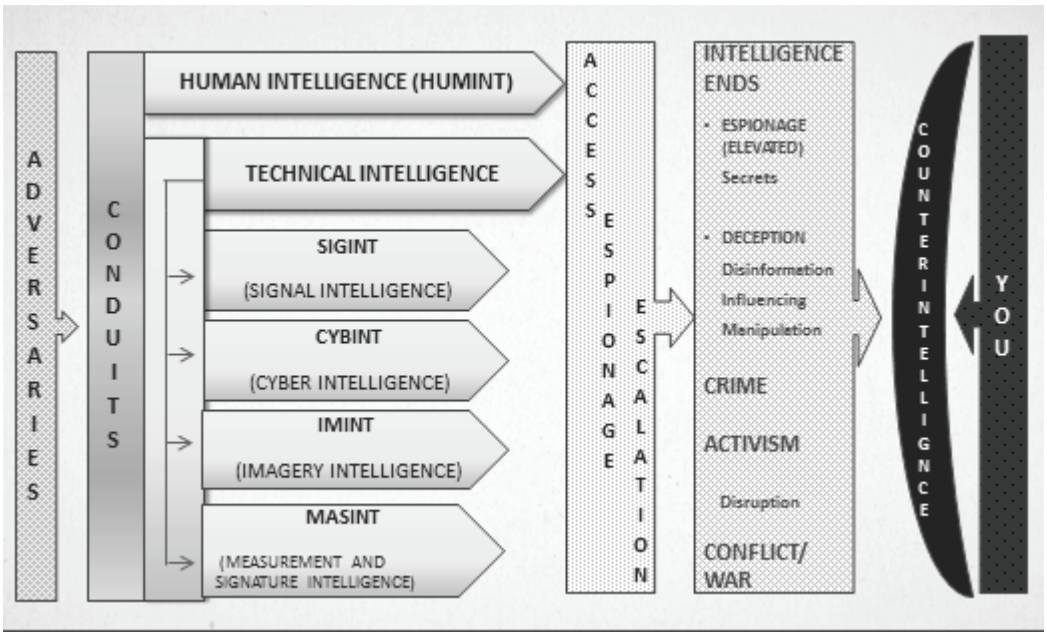

Figure 1: Adversarial intelligence: Conduits and ends

3.2 Counterintelligence: Measures, means and modes

To execute its mission, CI relies on measures and means that vary from the passive-defensive to active-offensive. At the one end of the spectrum, passive-defensive measures strive to deny adversaries access to protected information assets. They reduce vulnerabilities through a combination of policies, procedures and practices – referred to on a lighter note as "gates, guards, guns, and dogs" (Francq 2001). Apart from denying opponents access, properly instituted passive-defences measures are like caste walls. In addition to preventing common intrusions, their presence discourages intrusion attempts and consequently serves a deterrence function. Examples of passive-defensive measures are access and movement control, perimeter security, alarm systems, safes and vaults. At the other end of the spectrum, offensive counterintelligence aims to neutralise a competitor's intelligence efforts through measures ranging from deception and manipulation to the neutralisation of adversarial intelligence activities and systems (Duvenage 2013). Deception takes various forms and can be achieved through numerous means. Skilfully executed, deception attains a primary counterintelligence aim, which is the manipulation and control of an adversary. In the words of Codevilla (1992): "Action against the enemy through the enemy's own intelligence is the very consummation of CI."

Between passive-defensive and active-offensive lie a wide array of other measures such as pre-employment personnel security; in-service personnel security; technical surveillance countermeasures (TSCM); surveillance (physical, static, mobile and electronic); double agents; agents and continued monitoring. In most instances, these measures can serve defensive or/and offensive purposes. Defensive counterintelligence tactics and strategies provide information to, and act as triggers to alert the offensive side of the practice. Similarly, offensive operations (for example a source reporting on an adversary's intentions and capabilities) inform the pro-active configuration of defences. It will also be noted that several of these are highly useful in the collection of information on internal vulnerabilities (e.g. organisational weakness and the insider threat), the external environment as well as actual and potential adversaries. It goes without saying that without such information being analysed CI would be blind and unable to execute the defensive-offensive interplay. The following simplified matrix, per Figure 2, conveys CI's nuanced offensive-defensive interplay and the importance of collection:

DEFENSIVE MODE Denies adversaries access to and generates information about adversaries	
Passive Defence Denies the adversary access to information through security measures, systems and procedures.	Active Defence The collection of information on the adversary to determine its sponsor, modus operandi, network and targets. Methods include physical and electronic surveillance, dangles, double agents, moles and electronic tapping.
OFFENSIVE MODE Aims to manipulate, degrade, control and neutralise adversaries	
Passive Offensive Some forms of collection, disinformation and disruption. Include 'dummies' and 'decoys'.	Active Offensive The adversary is deceived and its interpretation manipulated. Disinformation can be channelled through for example double agents and 'moles'. Active-offensive CI could include some forms of collection and Covert Action. (Covert action aims to influence role-players, conditions and events without revealing the sponsor's identity.)

Figure 2: A four-sector counterintelligence matrix (compiled by the authors on the basis of narratives in Prunckun 2012, Sims 2009 and Godson 2001)

3.3 Counterintelligence process

The preceding matrix and discussion above, demonstrate CI as an intricate and exhaustive discipline. It is not only about defences, but also the concrete advancement of own interest vis-à-vis adversaries. No matter how well-resourced, the CI endeavour cannot protect all assets and advance all interests all the time. Prioritisation is critical. Since CI is not and end itself, optimal prioritisation requires CI to be at the centre of a business (or other entity's) strategy and Intelligence. While part of business Intelligence and strategy, CI is executed by means of a unique counterintelligence process which provide for the offensive and defensive thrusts. The traditional (positive) intelligence cycle, adaptations of the latter or other alternatives such as Clarke's target-centric process (Clarke 2004) simply do not work in CI – not conceptually and not in practice. Although it cannot be described within the confines of this paper, a CI process model proposed by Duvenage & Hough (2011) is proving useful also within CCI.

This section provided a primer that demarcated CI, explained CI measures and modes and reflected on the CI process. Building on this overview of CI, the paper proceeds with conceptualising CCI.

4. Conceptualising cyber counterintelligence

This section provides a provisional definition of CCI, advances a proposition for integrating CCI with CI and Intelligence and outlines some CCI methods, means and modes.

4.1 Defining and delineating cyber counterintelligence

While various definitions for CCI have been advanced, none of these specifically explicate the relationship between CCI and CI (c*f*. Carrol 2009, Bodmer et al 2012). In keeping with the paper's central contention, CCI is defined as that subset of multi-disciplinary CI aimed at deterring, preventing, degrading, exploiting and neutralisation adversarial attempts to collect, alter or in any other way breach the C-I-A of valued information assets through cyber means. Expanding on this definition, it is postulated that CI delineates CCI on the following three tiers (Duvenage & von Solms 2013):

- Applied to the cyber context, CI theory and practice provides a conceptual template for modelling CCI actions in the safeguarding and advancing of cyber interests. Mirroring CI, CCI has offensive and defensive missions that are distinguishable but not separable.
- To be effective, cyber counterintelligence needs to be interlocked with all-field counterintelligence – defensively and offensively. In this sense, CI cements an integrated approach to securing the cyber space. CCI is thus about both the modelling of cyber actions on CI, and the integration of these offensive and defensive actions with conventional CI.
- Effective CI protects and promotes the Intelligence endeavour and business strategy. Since CCI is part of CI, it is also integrated in business strategy and Intelligence.

Figure 3 depicts this three-tiered relationship graphically.

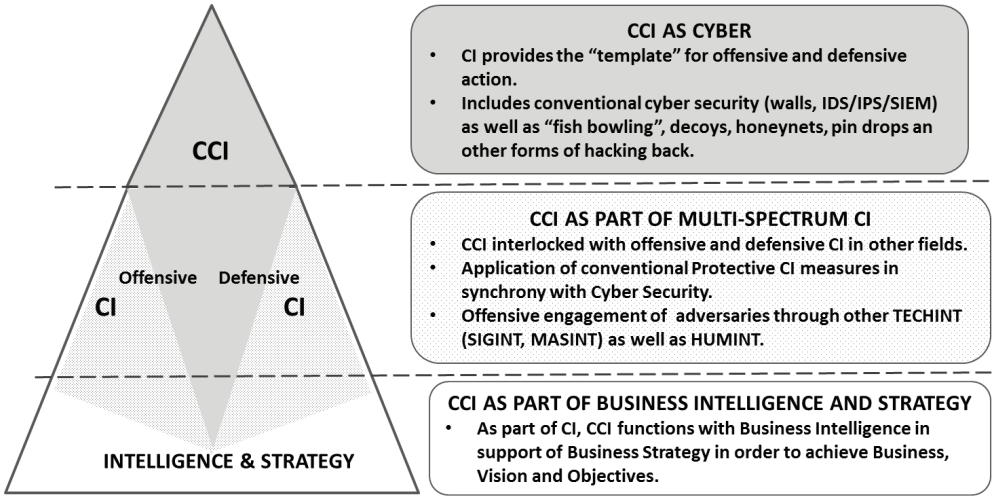
CCI
Offensive
CI
Defensive
CI
INTELLIGENCE & STRATEGY
CCI AS CYBER
• CI provides the "template" for offensive and defensive action.
• Includes conventional cyber security (walls, IDS/IPS/SIEM) as well as "fish bowling", decoys, honeynets, pin drops an other forms of hacking back.
CCI AS PART OF MULTI-SPECTRUM CI
• CCI interlocked with offensive and defensive CI in other fields.
• Application of conventional Protective CI measures in synchrony with Cyber Security.
• Offensive engagement of adversaries through other TECHINT (SIGINT, MASINT) as well as HUMINT.
CCI AS PART OF BUSINESS INTELLIGENCE AND STRATEGY
• As part of CI, CCI functions with Business Intelligence in support of Business Strategy in order to achieve Business, Vision and Objectives.

Figure 3: The cyber counterintelligence pyramid (created by the authors)

The postulation, per the narrative and Figure 3, is admittedly cursory and does not purport to conform to the criteria of a conceptual model. However, it could provide a useful premise for further research and the development of a conceptual model for implementation in the cyber domain.

4.2 Overview of cyber counterintelligence methods, means and modes

The section above made graphic and narrative reference to defensive and offensive CCI actions. Mirroring CI in general, CCI methods and means can be deployed in offensive and defensive modes, but defy categorisation in watertight compartments. At the very ends of this spectrum there are a few methods and means that could be designated clearly as active-offensive (notably cyber weapons with destructive purpose such as Stuxnet) or passive-defensive (e.g. access control and validation directives). In the main, however, offense-defensive and active-passive are not neat compartments, but rather the manner in, and end towards which, methods and means are deployed (Duvenage & von Solms 2013). This is illustrated in the following matrix, in Figure 4, which depicts the four cyber-counterintelligence modes (postures) an entity could assume:

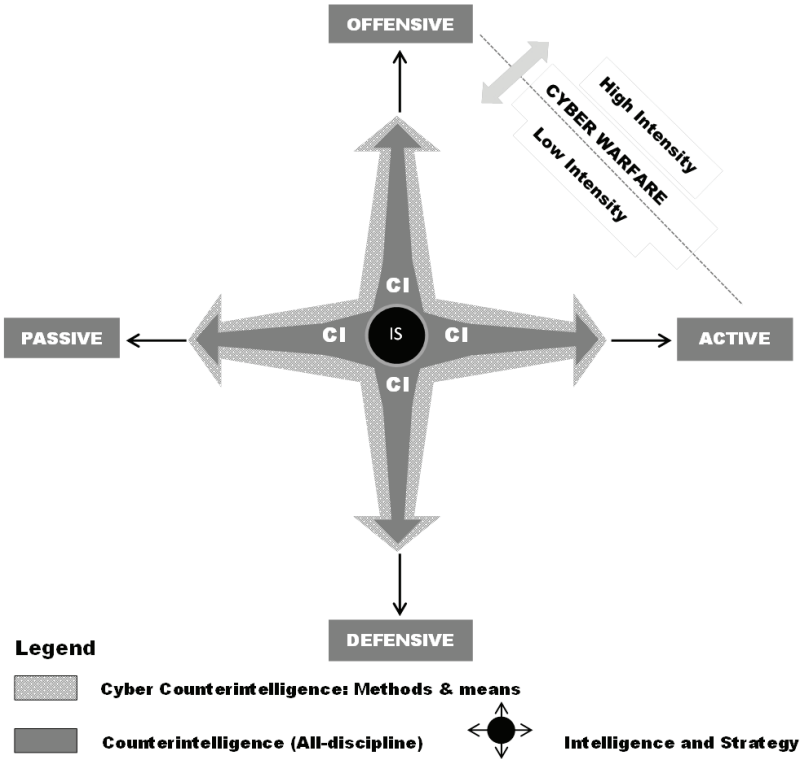

Figure 4: Cyber counterintelligence matrix (created by the authors)

The CCI matrix per Figure 4 is more than a notional construct and can be applied practically in the plotting of CCI methods and means. Utilising the matrix a presence is maintained, or at the very least that contingency planning is done in respect of all four quadrants. It furthermore facilitates innovation and creativity in the application of methods and means. Contrary to a misconception, for example, an Intrusion Prevention System can be configured with surprising positive results in executing aims in the other three quadrants. Consequently, the construction of a tabulated taxonomy of CCI methods and means could very well be an oversimplification. Nonetheless, at this early stage of conceptualising CCI such a simplification serves as a soundboard for further debate. With this caveat, a cursory taxonomy of CCI methods and means is provided per Table 1 (for reasons of layout featured on the next page).

Self-evidently, Table 1 samples only some of the wide array of CCI methods and means. Moreover, and given the length constraints of an article, on a very few of these are further elaborated upon, namely honeypots and decoys, cyber profiling and cyber-agent operations.

It would have noticed that of the means cited above, honeynets feature prominently in the active, passive as well as the defensive and offensive modes. Honeynets have been in use for more than two decades with the principle objective to detect, monitor and gain intelligence on adversarial intrusion on a network (Bodmer et al 2012). In recent years, the purposes of honeynets diversified from its original mostly defensive use to include also a much more active and/or offensive role. Concurrently, the different types of honeypots as well as configurations thereof, are sharply increasing. In respect of architecture, honeynets can be centralised, distributed, federated, and confederated (Bodmer et al 2012). The diversifying aims of honeynets now include one or a combination of deception, disinformation and the draining of adversarial resources through labyrinths and "rabbit holes" (Nakashima 2013, Duvenage & von Solms 2013). In a similar vein, decoys are highly useful in disrupting external intrusion and/or mitigating the insider threat (Voris et al 2013). The more resourced and sophisticated the adversary, the greater the imperative to attune the staging of honeynets and the content filling of honeypots, honeyfiles and honeytokens in accordance with the opposition's interests and modus operandi (Duvenage & von Solms 2013). Remark: This narrative of Section 4.2 continues after Table 1.

Table 1: Taxonomy of cyber counterintelligence methods and means (adopted from Duvenage & von Solms, 2013)

Defensive Mode		
Passive → ***Active***		
Deny → ***Detect*** → ***Collect***		
Physical Defensive	Personnel/User Defensive	System Defensive
Protects against: • Unauthorised access to facilities and systems. • *In loco* theft of data, hardware. • Introduction of malware through physical access to systems. • Physical destruction. • Unauthorised reading (acoustic, visual, analogue, signals). • While not conventionally seen as a Physical Defence, supply-chain management has a physical defensive function. It is also part of System Defences. Remark: Within the area of Physical Security, there is an extensive convergence of CCI and conventional CI. In keep with the article's central contention that CCI ought be	Consists of aspects such as • IT and user personnel **vetting**, re-vetting and confidentiality agreements. • Personnel security measures, **BYOD** user parameters or exclusions. • **User programmes** in cyber security (e.g. handling of security incidents and malfunctions). • Overlapping with system defences, the use of **software decoys** to mitigate the insider threat. • **Investigations** focused on cyber security incidents involving personnel. Could also include digital forensic investigations.	Comprises a combination of • Hardware and software such as ✓ Network perimeter-based security (filters, certain firewalls, etc). ✓ Malware scanners. ✓ Integrated automated systems/tools (that collect and evaluate information about devices connected to a network, activities thereon – inclusive of intrusions). Examples of such tools, discussed further on in the table, are decoys and honeynets. ✓ Overlapping with the latter, are **IDS** and **IPS.** Depending on its configuration, a honeynet can be defensive or offensive in type/mode. The term fish bowling denotes the defensive configuration. (Remark: See http://ids.cs.columbia.edu/content/ publications. html for extensive work that has been done on IDS/IPS). • Processes (such as supply- chain management, product verification and **testing**) are also in part system defences. • **Vulnerability assessments** and **penetration**

<table>
<tr><td rowspan="4">seamlessly integrated with CI, the sub-category 'Physical Defensive' is included in this taxonomy. Other classifications may exclude from CCI some of the measures listed above.</td><td colspan="2"></td><td>testing.
• Incident investigation and response. A CERT is per definition defensive – although it might contain offensive elements in its responsive action.
• BYOD regulation in as far as network interfacing is concerned (Also part of Personnel Defenses).</td></tr>
<tr><td colspan="3">• The use of honeynets and software decoys to mitigate the insider threat is an overlap between personnel and system defensive measures. They are mostly active CCI means.</td></tr>
<tr><td colspan="2">• Investigations focused on internal cyber security incidents involving personnel. May include digital forensic investigations.</td><td>• Investigations of external cyber intrusions could be part passive and part active system defence.</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">Offensive Mode</td></tr>
<tr><td colspan="4">Passive ⟶ Active</td></tr>
<tr><td colspan="4">• Collect → Disrupt → Exploit → Destroy</td></tr>
<tr><td colspan="2">• Collection of information on and the monitoring of the cyber sphere to detect cyber adversaries and their exploitation of the cyber sphere in a manner that is not own-network restricted (i.e. requires more than deployment of systems described under defensive mode). Could, depending on configuration also include IDS/IPS, honey-client applications (as opposed to host-based honeypots) and data mining.
• The recruitment and handling of virtual agents on underground forums (under true or false flag) that can serve the purpose of collection and/or exploitation. (Under certain circumstances virtual agents can also develop into HUMINT assets).</td><td>Measures taken to exploit and neutralise adversaries activities in the cyber sphere:
• System and honeynet configured offensively with the aim of exploiting and deceiving adversaries. False information is displayed to adversarial reconnaissance tools, network scanners and listeners, etc. This has as one of its aims to lead adversaries in the direction of your own preference.
• Utilisation of virtual agents for offensive purposes.</td><td rowspan="2">Cyber warfare, in the full extent of the term, is typically excluded from the mandate of civilian intelligence communities. Nevertheless, a civilian intelligence CCI outfit would have the capacity for selective operations that have cyber warfare characteristics.</td></tr>
<tr><td colspan="2">• Cyberespionage on adversaries. Distinguishable from own-system collection (IPS, IDS, honeynets) on the basis that adversarial networks are actively targeted and exploited in accordance with strategic and operational objectives.</td></tr>
</table>

Counter-action with matching sophistication, in turn, requires sound analysis of high-grade information on the environment and adversaries. Unsurprisingly cyber profiling which involves the application of criminal and intelligence profiling methods to the cyber realm is fast gaining field as a CCI specialisation area (Bodmer et al 2012). In order to procure information on actual and potential adversaries, as well as to keep to tabs on hacking communities of all sorts, CCI outfits maintain a layered presence on nets and forums. This presence varies from the deployment of soft and hardware instrumentalities to the cyber equivalent of HUMINT counterespionage, namely the recruiting, turning and handling witting/unwitting agents (Duvenage & von Solms 2013).

4.3 Cyber counterintelligence as a multi-disciplinary subset of counterintelligence

In line with the theoretical outline of the relationship between CCI and CCI (Figure 3 and Figure 4), the practical safeguarding and advancement of cyber interests is a multi-disciplinary endeavour. CCI is thus not only multidisciplinary in itself but is overlaid upon multi-disciplinary counterintelligence. As part of the Edward Snowden revelations it was, for example, reported that the USA and UK Intelligence communities rely on the recruitment and running of HUMINT sources networks in the global telecommunications industry to "tackle" some of their "most challenging targets" - inter alia in the cryptology field (Ball et al 2013). In keep with such multi-dimensional threats, a CI operation in the cyber field could entail a multi-disciplinary team comprising

cyber security specialists, strategic analysts, tactical and technical analysts, HUMINT specialists (e.g. agent handlers and intelligence psychologists), cyber-defense technical experts, language experts, ethical hackers, sociologists and religious experts (Bardin 2011). While a sharp edge on the offense, humans are also the weakest and possibly the most ruinous chink in the defensive armour. Powell et al (2013) asserts "an organizations insiders" as "primary threats to cybersecurity … [which are] ….the most difficult to mitigate." Complementary to technical defences, CI personnel fidelity measures and HUMINT counterespionage practices are thus critical. This is being highlighted by unfolding detail around the Edward Snowden incident.

The convergence of cyber and HUMINT counterintelligence was furthermore demonstrated by a recent re-evaluation of the Aurora attacks. This re-evaluation suggests the Aurora attacks were not, as was initially thought, a People Republic of China (PRC) operation targeting human-rights activists. It was in fact a Chinese counterintelligence operation to determine whether PRC intelligence operations and agents have been compromised to USA intelligence (Corbin 2013). Duvenage & von Solms (2013) cite as a further example of "an integrated CI initiative, a disinformation campaign as part of which the staging and content filling of a honeynet is harmonised with disinformation fed to an adversary through a HUMINT asset (e.g. double agent)."

4.4 Cyber counterintelligence and counterintelligence – an integral part of intelligence and strategy

CCI, to re-state the paper's recurring theme, forms part of and is guided by the integrated CI endeavour. Consequently, CCI follows the CI processes noted in Section 3.3. CI not only safeguards a business, but renders inside information on competitors highly useful to executives. In addition; deception, disinformation and other such projects support a company in achieving its business objectives. The latter illustrates the theoretical postulations per Figures 3 and 4 which put Business Intelligence and strategy as the pivot around which CI and CCI evolves.

5. Conclusion

This paper forms part of the still spare yet fast growing body of academic literature which views CCI as a practicable approach for governments, businesses and other sizable entities for securing and advancing cyber interests. Proliferating threats and trends affecting cyber security are not all bad news. Contradictory as it may appear, the more extensive adversarial cyber action the greater the potential opportunity could be for counter-exploitation. The call for cyber CCI should not be misconstrued as a call for a free-for-all cyber Wild West. Performed haphazardly and in a silo, CCI could be self-destructive.

There are several pre-conditions for effective CCI. To be effective, CCI should be an integral part of multi-disciplinary CI – conceptually and in practice. In consulted academic literature, however, such conceptualisation is lacking. For the most part they endeavour to progress with CCI theory construction, without a sound foundational explication of CI. Theory so formulated and models so constructed could hold serious negative repercussions on a practical level. Within counterintelligence, the price for bad theory is costly failures. As pointed out in an earlier contribution: "Conceptual models are not mere theoretical, academic constructs. Models condition our thinking and our approach to practice. What we therefore need is a sound overarching CCI model that can synergistically bind developing theory." (Duvenage & von Solms 2013)

Therefore, this paper firstly aimed to put the counterintelligence in cyber counterintelligence. This was done through conceptualising CCI as part of multi-disciplinary CI and the applications of time-tested CI constructs to the cyber sphere. Secondly, the article offered a few conceptual constructs as contours towards the construction of such a model. So doing, it demonstrated the degree to which conventional, time-tested CI constructs can guide CCI's conceptualisation. The actual construction of a credible model, however, will require extensive in-depth, multi-disciplinary research and debate.

References

Ball, J. et al. (2013) "Revealed: how US and UK spy agencies defeat internet privacy and security" [online], *The Guardian.* September 06, http://www.theguardian.com/world/2013/sep/05/nsa-gchq-encryption-codes-security.

Bardin, J. (2011) "Ten Commandments of Cyber Counterintelligence", Adapted from Olsen, J. M. *"Ten Commandments of Counterintelligence"*, n.d.) *CSO Risk*. Posted on blog 2011-06-21. Retrieved on January 09, 2013 from of_cyber_counterintelligence_adapted_from_james_m_olson.

Clarke, R. M. (2004) *Intelligence Analysis: A Target-centric Approach*, CQ Press, Washinton D.C.

Bodmer, S. A. et al (2012) *Reverse Deception – Organized Cyber Threat Counter- Exploitation*, McGraw-Hill, New York.

Bodmer, S. A. et al (2014) *Hacking Back: Offensive Cyber Counterintelligence*, McGraw-Hill, New York.

Carrol, J. (2009) "Cyber Counter Intelligence", *Defense Tech*. Retrieved on December 03, 2012 from http://defensetech.org/2009/03/09/counter-cyber-intelligence/

Corbin, K. (2013) " 'Aurora' Cyber Attackers were really running Counter-Intelligence", *CIO*, Retrieved May 01, 2013 from http://www.cio.com/article/732122/_Aurora_Cyber_Attackers_Were_Really_Running_Counter_Intelligence?page=1&taxonomyId=3133

Codevilla, A. (1992). *Informing Statecraft – Intelligence for A New Century*. The Free Press, New York.

Dictionary.com (2014), *"counterintelligence"*, available at http://dictionary.reference.com/browse/ counterintelligence

Duvenage, P. C. (2010). *Open-source Environmental Scanning and Risk Assessment in the Statutory Counterespionage Milieu*. Unpublished doctoral thesis. Pretoria, University of Pretoria.

Duvenage, P.C. (2013) "Counterintelligence," in Prunckun, H. (ed.), *Intelligence and Private Investigation: Developing Sophisticated Methods for Conducting Inquiries* (Springfield IL: Charles C. Thomas, 2013).

Duvenage, P.C. and Hough, M. (2011) "The Conceptual Structuring of the Intelligence and the Counterintelligence Processes: Enduring Holy Grails or Crumbling Axioms – Quo Vadis?". *Strategic Review for Southern Africa*, University of Pretoria, Pretoria.

Duvenage, P. C. and von Solms. S.H. (2013). "The Case for Cyber Counterintelligence" Paper read at *5TH Workshop on ICT Uses In Warfare and the Safeguarding of Peace (IWSP'13)*, Pretoria, South Africa. November.

Francq, A. (2001). "The Use of Counterintelligence, Security, and Countermeasures", in Fleisher F.S and Blenkhorn D. L (eds), *Managing Frontiers in Competitive Intelligence*, Quorum Books: Westport.

Giles, L. (2002) *Sun Tzu – The Art of War,* (Translation), Dover Publications, New York.

Godson, R. (2001) *Dirty tricks or trump cards - U.S. covert action and counterintelligence*. New Brunswick, Transaction Publishers.

Helton, D. (2013) "Human Threat and Cyber Counterintelligence – an Agent's Perspective", *Speartip*, November 26. Retrieved from http//www.speartip.com/ on January 04, 2014./

IBM (2013) *IBM Protects Clients from Security Attacks with New Cloud Solution,* 22 Oct 2013, [online] http://www-03.ibm.com/press/us/en/pressrelease/42269.wss

ISC^2 (2013) Michigan Chapter. *Events Diary*. November 8, 2013, [online], http://isc2chapter-westmi.org/category/events/

Kuusisto, R. and Kurkinen, E. eds. (2013) Proceedings of the12th European Conference on Information Warfare and Security, Jyväskylä (Finland), Academic Conferences and Publishing International Limited, Reading (U.K).

Lües, J. (2012). "IT security has failed – Once effective, It security is now deteriorating on all fronts." iWeek, Issue 225, June 06.

Miller, N.S. (1980) "What is Counterintelligence – Discussants." in Godson R. ed. *Intelligence Requirements for the 1980's: Counterintelligence*. National Strategic Information Center, Washington (D.C.).

Nakashima, E. (2013), "To thwart hackers, firms salting there servers with fake data", The *Washington Post*, January 03.

Park, Y. and Stolfo, S. J. (2012). Software decoys for Insider Threat. Seoul. ASIACCS. http://ids.cs.columbia.edu/content/publications.html

Powell, D, Wick, A and Fergus, D. (2013) "Protecting against Cyber Threats" *Security Management*, (online), retrieved on May 11, 2013 from http://www.securitymanagement.com.

Prunckun, H (2012). *Counterintelligence: Theory and Practice*. Plymouth, Rowman & Little Publishers.

Sims, J. E. (2009) "Twenty-first-Century Counterintelligence" in Sims, J. E. and Gerber, B. (eds.) *Vaults, Mirrors and Masks – Rediscovering U.S. Counterintelligence*. Georgetown University Press, Washington (D.C.)

Van Cleave, M. K. (2007) *Counterintelligence and National Strategy*. School for National Security Executive Education. Washington (D.C.): National Defense University Press. Retrieved on July 02, 2010 from www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA471485

Voris, J et al. (2013) *Bait and Snitch - Defending Computer Systems with Decoys*. Columbia University, New York.

World Economic Forum (2014) *Insight report - Global Risks 2014,*(online), www.weforum.org Global , WEF, Geneva.

Information Security Economics: Induced Risks and Latent Costs

Evangelos Frangopoulos[1], Mariki Eloff[2] and Lucas Venter[3]
[1]School of Computing, UNISA, Pretoria, South Africa
[2]Institute for Corporate Citizenship, College of Economic and Management Sciences, UNISA, Pretoria, South Africa
[3]Institutional Director: Research Support and Extraordinary Professor: Computer Science and Information Systems, North-West University, Potchefstroom, South Africa
vfrangopoulos@hol.gr
eloffmm@unisa.ac.za
lucas.venter@nwu.ac.za

Abstract: Even though the direct effects of financial issues may seem obvious at first glance as far as the efficiency of Information Security measures is concerned, there are also many more indirect and subtle ways that Economics affect and are affected by the effort towards attaining the desired levels of Information Security. As such, the Economics of Information Security should constitute a significant part of the Information Security Posture Assessment process and be directly addressed in this context. As the complexity and interdependency of Information Systems augments and new technologies lead to the de-materialisation of Information Systems assets, it becomes progressively evident that the conflicting interests and incentives of the various stakeholders involved in the operation of an Information System affect the overall Information Security Posture of the entire system, perhaps even more significantly so than technical or policy limitations. This paper examines economic considerations from an Information Systems Security / Cybersecurity viewpoint and sets the ground for further work that will integrate economics, along with technical and non-technical issues, into the Information Security Posture Assessment process.

Keywords: cybersecurity, economic considerations, risk management, infosec economy

1. Introduction

It is a generally accepted truth that Information Security (IS) costs money. There are inherent costs in all aspects involved in the mitigation of the risks relevant to IS, such as the technology used, required procedures, statutory compliance, awareness education etc. The same holds true for Information System Security, or Cybersecurity, which is the IS subset that deals with Information Systems aspects of IS (what used to be known as "computer security"). The expenditure for IS has been steadily increasing over the past years, following the increase in emerging Cybersecurity threats as it has been established in the recently released "Global State of Information Security® Survey for 2014" (PwC, 2013a). The "Emerging Threats" vs. "Cybersecurity initiatives" race is being viewed as the modern cold war face-off, which is constantly fuelled by large numbers of emerging threats as categorically shown in the ENISA Threat Landscape 2013 report (ENISA, 2013). Unfortunately, it may well be that even though IS expenditure increases, it does not do so at a rate capable of efficiently counteracting the emerging threats. Furthermore, the attackers are no longer lone hackers driven by personal agendas, but there may well be criminal organisations or even nations behind cyber-offensives.

The ugly truth is that this "war" is definitely asymmetric, as the attackers have many weapons in their arsenal and attacks can be mounted with low cost, compared to the possible gain stemming from a successful attack. On the contrary, regarding the defensive side, any "Cybersecurity initiative" faces the real risk of becoming a bottomless money-pit or, if funding is prematurely curtailed, not being able to justify its existence, and, even worse, provide a false sense of security to the defending force.

In the preface of their book titled "Foundations of Information Security: Based on ISO27001 & ISO27002", Hintzbergen et al. (2010) make reference to the Great Wall of China as the ultimate example of the effort towards physical security and the effect of physical security in modern-day IS. The Great Wall of China, and fortress walls in general, are indeed very popular examples used in the IS literature (Groom, 2003) from which analogies are drawn (think of Bastion hosts and Gateway servers). However, the Great Wall of China can also be used to demonstrate the asymmetry of cyberwar. The Great Wall was built at great expense and effort, yet it too had vulnerable points, partly due to terrain and partly due to the limitations in its construction. A single well-targeted attack against the wall could result in a breach, and thus intrusion, with a total cost in effort and resources which was negligible compared to the total cost of construction and maintenance of the Great Wall.

It is thus not unfitting to draw analogies from this example for the Cybersecurity paradigm, whereby the equivalent “Great Wall” is constructed by a plethora of technical measures, supported by policies and procedures. The problem, though, frequently lies in the “brick and mortar” of this construct and manifests itself in the form of hastily produced and inadequately debugged software that, when integrated in the wall-building effort, comes with inherent defects and weaknesses. This is usually caused by the desire of software houses to minimise the time-to-market of a new product and maximise the profit on the relevant investment. Other problems with the “construction material” of the “wall' may result from choices made by the defending champions regarding the selection and integration of security-crucial components which are, once more, influenced by financial considerations. Furthermore, insofar policies and procedures are concerned, they, too, are governed by direct financial decisions, as well as by indirect ones, such as those that are related to productivity, financial risk, reputation issues etc.

2. Vulnerabilities induced by economic considerations

Economic considerations are being used in the process of risk assessment, usually to define an organisation's “risk appetite” and lead to informed decisions regarding the implementation of security measures. Even though there is nothing intrinsically wrong with this procedure, which does indeed produce valid results, perhaps a change of paradigm is in order: the economic considerations themselves should form part of the threat landscape that needs to be examined in the context of a risk assessment.

In such an approach, economic considerations will not only function as the yardstick against which every type of risk and its compensatory measures are benchmarked, but they, too, will come under scrutiny and evaluation.

In the following sections, an effort is made to bring to light a number of economic considerations that give rise to vulnerabilities or may even be themselves identified as vulnerabilities for IS.

The question that needs to be addressed is whether, and to what extent, the economic considerations that may seem justified from the Management's point-of-view, adversely affect IS, and Cybersecurity in particular, and if these can be somehow controlled.

2.1 Diverging incentives and holes-in-the-wall

It was already mentioned that the defensive wall that organisations are trying to build around their information assets may be undermined by the different incentives between the organisation and the software and hardware vendors. One such difference may indeed be the fact that software vendors are commonly more concerned with the production of buyer-acceptable software at minimum cost and with minimal time-to-market, rather than with the effort towards the security of their product (Anderson & Moore, 2006) which is usually exhausted at the lowest obligatory level prescribed by regulatory compliance (Shostack & Stewart, 2008, p.107). It is interesting that at this day and age software houses still shy away from assuming any of the risk stemming from the use of their products (Marotta-Wurgler, 2007 & 2011). In order to use the software, clients have to accept disclaimers that minimise or even eliminate the liability of the software producer/vendor. Effectively, as Bessey et al. (2010) put it: “Indifference can arise from lack of accountability”.

However, even though users may like the idea of secure software, they can neither really tell the difference between secure and insecure software nor are they inclined to pay for better-written software. Thus, because of competition, the costly effort that goes into the production of secure software cannot be adequately compensated for (Anderson & Moore, 2006). This, in turn, makes it difficult for software houses to allocate funding and resources for the production of secure software. If software security were somehow enforced, software would probably cost significantly more than it does today.

The problem of diverging incentives between the software houses and their ultimate clients becomes even more serious if stories, such as the one reported by Reuters (2013) exposing a secret $10 million contract between RSA and NSA, prove to be true. If products by Information Security giants that are used by hundreds of software producers and service providers are tainted in the manner reported by Reuters, then innumerable end-users ranging from individuals to corporate entities, to governmental and international organisations, are adversely and grossly affected.

2.2 Upgrading towards greater risk

The function of Information Systems in a constantly changing landscape of emerging threats requires systematic updates in order to maintain an acceptable level of Cybersecurity. Continuous updating is in part necessitated by initial inadequate debugging of software code. Hence, updates become tools for fixing bugs and addressing security liabilities during the production phase of software (Brown, 2005). Complete debugging becomes virtually impossible given the complexity of the code and the constraints imposed on software production by economic considerations. It is also assumed that once a "patch" is applied in order to control an existing short-coming in the code or provide new functionality, the additional code may itself have new and different short-comings that create new vulnerabilities to be addressed and so on. However, it is reasonable to assume that with every iteration of the update process, the software tends asymptotically to a state of higher security that should eventually -ceteris paribus- lead to a higher level of Cybersecurity. Unfortunately, all other things do not remain constant, and "new and improved" versions of the software, nomically give rise to a myriad of new vulnerabilities.

In practice, even if not required or requested by the customer, the upgrade to the new version of the software is compulsory -usually at significant cost-, as support for previous versions is discontinued eventually and thus they become obsolete. The new versions, in order to provide increased functionality that appeals to the customer, are always more complex, with more lines of code and thus include more bugs, effectively feeding the vicious cycle of the continual struggle towards Cybersecurity. The paradox of this situation is that the end-user invariably uses only a small fraction of the software's full functionality. By trying to address the differing individual needs of their clientèle all at once, the software producers create coding behemoths that are impossible to contain from a security perspective. The end result is that the customers are forced to accept increasingly expensive and at the same time more vulnerable software that requires a larger number of more expensive controls. Hence, software becomes more expensive & less secure.

2.3 Productivity vs. security

Information Systems Security / Cybersecurity measures are often seen as inhibitors of productivity. This may manifest itself at the user level where, for example, USB port deactivation or the necessity for many and complex passwords may cost time and effort on the part of the user. Hence, the user may feel that this is a waste of resources that reduces his productivity and eventually the profits of the business. When this view is shared by the management (after all, managers are usually also users of the same Information System), then:

a) at a critical time they might directly or indirectly force the users or even IT administrators to break away from the Information Security Policy (ISP) and circumvent the controls in order "to get the work done" (the image of a CEO wielding a tablet and asking IT people to connect it to the network despite ISP provisions, does spring to mind), or

b) oppose the inclusion of controls that might feel counter-productive -and thus financially unwise- in the ISP, effectively undermining the Cybersecurity effort.

This is corroborated by the findings of Rainer et al. (2007) where the list of top issues that Information Security professionals have difficulty addressing with management includes "top management support", "low funding and inadequate budgets" and "justifying security expenditures".

2.4 Moving away from the "moat & castle" cybersecurity model

Pfleeger (1997) used a medieval analogy of a castle surrounded by a moat in order to explain the idea of "defence-in-depth" in Information System Security. This has been and still is the model through which most information assets are protected. However, in the interest of enhanced productivity, which directly translates to profit maximisation, business management continually requests greater availability of information assets, irrespective of the geographic location of users and time-of-day. From low-end IT users to business managers we are all enticed by the flexibility and mobility that technological solutions provide for our everyday life and we are willing to sacrifice large parts of our privacy and right to confidentiality in order to keep enjoying this maximised availability of information. Being accepted at a personal level, this model is then unreservedly requested for the management of professional information by turning a blind eye to the extended security requirements that apply to this type of information. In order to provide such services and facilitate relatively

new and "financially sensible" notions such as "Bring Your Own Device" (BYOD) and the use of the Cloud, Information Systems have to move away from the Pfleeger "Moat & Castle" paradigm and, in the attempt to remain secure, confront a huge new set of threats. It could be argued that this is evolution and progress and that it should not be hindered but, instead, adaptation is in order. However, recent history has shown that the move towards new technologies is anything but systematic and controlled as far as the security of these new technological solutions is concerned (GIT, 2013). On the contrary, technical solutions are proposed and adopted on short-term and immediately visible economic merits -or the expectation thereof-, before the degree of their security resilience is even established, leading to great loss, financial and otherwise, during their maturation. Typically, even though we are being taught that security considerations for a proposed solution should begin at the time of its conception, the fact remains that most of the IT industry generally works under the principle of putting a new product out for sale as quickly as possible and worrying about its security parameters later. Thus, we find ourselves stepping outside the relatively secure perimeter defined by the "Moat & Castle" in order to gather larger crops, without having first ensured that we have a fighting chance against an enemy attack.

2.5 To be or not to be (secure)?

In the various certification programs for Information Security professionals (as in ISACA, 2011) it is taught that the effort towards Information Security should function as a business enabler by a) being aligned with business goals and objectives, b) ensuring regulatory compliance and c) reducing risk to an acceptable level. These are valid working hypotheses, as, indeed, overspending on Information Security would be unwise and would have adverse effects on the "bottom line" of the business. In this sense, several spending models and investment approaches for Information Security have been developed (as in Gordon & Loeb, 2002; Cavusoglu et al., 2004; Tsiakis & Stephanides, 2005; Böhme, 2010 etc). It is beyond the scope of this paper to compare and analyse Information Security spending models and a very extensive bibliography exist on the subject. However, despite the outcome of a model-based expenditure analysis -with all of its merits and shortcomings- the ultimate business decision on information security must be aligned with the core business objective and it must be made by people whose roles are closer to management than to Information Security. Schroeder and Grimaila (2006) showed that there is definite bias in the decision-making process regarding Information Security and that the decision-makers "will place more weight on operational outcomes than security outcomes". Thus, even if security considerations are against the operation of a system, the decision-makers are perfectly capable of ignoring that outcome, simply because the business impact of not having the system in operation is greater than allowing it to function without sufficient security. If such economics-based decisions are abused, they will significantly raise the organisation's risk appetite and undermine its Cybersecurity posture.

2.6 Regulatory provisions as an excuse

Apart from business alignment and acceptable risk as examined above, there is also the issue of regulatory compliance. It is evident that defensive action in Cybersecurity always comes as a result of an exploited vulnerability. Strictly speaking, pro-active cyberdefence -in the exact meaning of the term- is not really possible. Most cyberdefence measures are thus one step behind the respective threats in this continual tug-o-war. To make matters worse, for such a measure to be included in a regulatory compliance document, even more time has to pass. It is thus only reasonable to conclude that regulatory provisions only cater for "yesterday's problems". This is no secret among serious Information Security professionals, and thus regulatory compliance is, by default, considered inadequate in real Information Security and Cybersecurity terms. On the other hand, regulatory compliance gives a golden excuse to Management to limit Information Security expenditure because it simply allows the tick-off of another checkbox in a financially justified "to do" list, without effectively addressing the underlying security issue (Shostack & Stewart, 2008, pp.35-36).

2.7 A cloudy future

While the move towards the cloud does seem to have some obvious financial advantages, the cloud still remains a very unsafe place for storing information (GIT, 2013). Furthermore it constitutes an example of how large IT vendors use their market penetration to further shape the market (and its security) according to their interests. The degree of success of a Cloud service provider depends on the provider's ability to create service capacity at a much larger scale than any individual organisation could manage and then sell it back to the user piecewise at a gain, but at a much lower cost than the user could manage alone (Kushida et al., 2011). In order to quickly attract customers, the major cloud service providers created a very attractive but quite unsafe

environment. According to the Georgia Institute of Technology "Emerging Cyber Threats Report 2014" (GIT, 2013), for data to be moved to the cloud there must be trade-offs between security and usability as "File sharing and other cloud services still have questionable security". The report identifies three key issues that despite all predictions to the contrary, persist: a) Business data is moved to the cloud protected only by the security measures provided by the cloud storage firm, b) private-key encryption is not used, as storing encrypted data in the cloud drastically reduces the cloud's utility, c) the use of searchable encryption necessitates trade-offs between security, functionality, and efficiency. Yet, even though security is all but non-existent in the cloud, companies and organisations exhibit a huge appetite for risk by deciding to keep using it, obviously basing their decision on financial merits alone. Even hardware methods of controlling who has access to cloud data and where data is stored, through the use of Trusted Platform Modules or "TPM" as described by Krauss et al. (2013), may be insufficient, as even these most secure Integrated Circuits (ICs) have been reported to succumb to physical attacks using acid, rust remover, a lot of time and a lot of skill (Tarnovsky, 2008). As TPM ICs are used in many sensitive applications such as secure communications, military systems and the like, it is conceivable that if enough money and resources are allocated to the effort, a great deal of secrets can be had; once again, an economic dilemma.

2.8 Cutting chip corners

Since the 1980's, the hunt for lower production cost and profit margin maximisation on digital equipment has led to a shift of Integrated Circuit foundries, from their birthplace in Silicon Valley to the Far East (Perera, 2012). This has resulted in most of the big names in IC design becoming "fabless" producers, i.e. producers without fabrication facilities of their own, who subcontract the production of their designs to independent foundries, mostly in the Far East. According to the Solid State Technology Insights website (SST, 2012), foundries in Taiwan, South Korea and China occupied 9 of the 12 top worldwide spots of semiconductor foundries for 2012 (the US are still present on the list with two foundries and Israel with one). For China in particular, according to PwC's report on "China's impact on the semiconductor industry 2013 update" published in September 2013 (PwC, 2013b), China's semiconductor consumption market grew by 8.7% in 2012 to reach a new record of 52.5% of the global market, while its share of worldwide electronic equipment production increased to 34.2% in 2012 and is expected to increase to more than 40% by 2017. Accordingly, China's share of worldwide semiconductor production reached 12% in 2012 and is expected to reach at least 14% by 2015. The reliance of the world electronics industry on far-eastern semiconductors, a choice made on the grounds of lowering the cost of production, appears to have a darker side: for many years there have been rumors that Integrated Circuits may contain "undocumented features" ranging from "kill switches" to backdoors. One of the first sensational stories about a kill switch had to do with the temporary "blindness" of Syrian radars when Israeli bombers carried out a raid in 2007 against Syrian targets (Adee, 2008). Compromised ICs in the radar systems, containing a "kill switch" were blamed at the time and theories about creating hardware backdoors on ICs followed. Recently, Skorobogatov and Woods (2012) provided an end to the rumors by actually locating a backdoor on a military microchip fabricated by an independent foundry for a fabless IC producer. Given the proliferation of personal mobile devices and the evolution of "Internet of things" which is already well under way, the implications of the use of ICs with malicious payload on Cybersecurity are obvious. The risk increases as tools like the "Shodan" engine (www.shodanhq.com) which searches for exposed devices on the Internet, are becoming available. Unless procedures for secure and authenticated manufacture of ICs, as well as methods for full post-production verification of ICs are devised (something which at this point proves to be very difficult), hundreds of millions of individuals, corporations and organisations around the world will face unprecedented risks. However, the cost for the required procedures and methods will be significant and it remains to be seen if the producers of equipment will decide to accept it, or prefer the creation of new vulnerabilities for the sake of the "bottom line".

3. Proposed controls

The presented cases only serve as examples of how considerations of an economic nature may directly or indirectly introduce vulnerabilities that undermine the effort towards Cybersecurity. The obvious question is whether effective controls can be devised.

As far as bad coding is concerned, this could be addressed through compulsory compliance to improved software industry standards that enforce secure coding practices such as the ones presented by CERT (2011) and OWASP (2010). As the infrastructure for assessing software structural quality and non-functional requirements, based on code metrics, does exist (e.g. standards ISO/IEC 25010:2011 and 9126-1:2001), it is

only a matter of creating mechanisms that ensure code compliance to secure coding standards or improving existing ones (Devanbu & Stubblebine, 2000; Chen & Wagner, 2002; Baggen et al., 2012). A widely-recognised seal of approval could then be given to security-compliant software.

For this idea to bear fruit it will be necessary for major software houses to support it rather than oppose it because of the short-term financial burden that it might cause.

A change of mentality might be in order insofar users' expectations from software are concerned. Instead of using complex and thus security-challenged software, a decision should be made to use simpler software without superfluous functionality but with fewer security deficiencies. Software should be tuned to an organisation's needs following an accurate requirements' analysis. This will not lead to cheaper software but it will help allocate funding more prudently towards increased, and distributed, Cybersecurity. Modular programming could help towards this end, as presented, amongst others, by Chen & Wagner (2002), Bauer et al. (2003), Dhiman et al. (2013).

To aid in this direction, markets for vulnerabilities (created by companies who purchase zero-day vulnerability information), can be used to quantify software security, thereby rewarding good programming practices and punishing bad ones as proposed by Anderson & Moore (2006).

In the above context, if insurance against cyberattacks matures, it will be helpful through the provision of metrics for quantifying risks induced by poorly-coded and security-challenged software. Additionally, accurate evaluation of the cyber-risk levels of organisations will act as an incentive towards better Cybersecurity. Organisations will be keener to adopt cyberdefence practice as this will reduce their Cybersecurity insurance premiums, leading to a distribution and normalisation of the Cybersecurity effort in a manner similar to which vaccinations help eradicate disease: they work only if, ideally, everybody participates.

Stricter regulatory compliance should be in place so that when software houses, service providers or infrastructure producers are found to have been knowingly working against the interests of their customers/end-users, the loss suffered by the users should be transferred in whole or in part to the offending party. Given the de-materialisation of Information Systems assets, this can only be made possible through the application of international law.

Finally, the gravity of the Cybersecurity component in the decision-making business process should be stressed. IS/Cybersecurity experts cannot be management experts and vice-versa. For this reason, security steering committees should comprise experts from all disciplines and Cybersecurity issues be assigned the importance they deserve and even be given priority over “bottom-line” financial considerations. To accomplish this, awareness programs must be more frequent and their effect be monitored. From the authors' experience, once decision-makers internalise the true gravity of Cybersecurity issues, they tend to heed the advice of IS experts more closely.

4. Conclusions and further work

A non-exhaustive list of representative examples of economic considerations that lead to IS vulnerabilities has been examined. Controls for vulnerabilities directly or indirectly caused by economic considerations are of a financial nature themselves and where that is not effective, institutional compliance must come into play. New technologies should be allowed to mature and pass through the crucible of time before being used in IS-critical situations. Technologies used for inconsequential communication and data management should not be ported to an IS-critical environment. Software houses may have to be institutionally forced to be more honest in their marketing habits and their customers must learn to exercise self-restrain and not be gluttonous in their IT-related desires. Users and, most importantly, decision-makers must be re-educated on the value of privacy and security in cyberspace and the collective awareness level on IS and Cybersecurity must be raised.

The next step should be to systematically introduce economic considerations as vulnerability inducers in the IS/Cybersecurity posture assessment process.

References

Anderson, R. & Moore, T. (2006) The Economics of Information Security. *Science*, 314, pp.610-613, 27 October 2006.

Adee, S. (2008) The Hunt For The Kill Switch. *Spectrum*, IEEE, 45(5), pp.34-39, May 2008.

Baggen, R. et al. (2012) Standardized code quality benchmarking for improving software maintainability. *Software Quality Journal*, 20(2), pp.287–307.

Bauer, L., Appel, A.W. & Felten, E.W. (2003) Mechanisms for secure modular programming in Java. *Software: Practice and Experience*, 33, pp.461–480.

Bessey, A. et al. (2010) A few billion lines of code later: using static analysis to find bugs in the real world. *Communications of the ACM*, 53(2), pp.66–75.

Böhme, R. (2010) Security metrics and security investment models. In *Advances in Information and Computer Security*. Springer, pp.10–24.

Brown, D.J. (2005) An update on software updates. *Queue*, 3(2), pp.10–11.

Cavusoglu, H., Mishra, B. & Raghunathan, S. (2004) A model for evaluating IT security investments. *Communications of the ACM*, 47(7), pp.87–92.

CERT (2011) *CERT Top 10 Secure Coding Practices*. [online] Available at: <https://www.securecoding.cert.org/confluence/display/seccode/Top+10+Secure+Coding+Practices> [Accessed 18.1.2014]

Chen, H., & Wagner, D. (2002) MOPS: an infrastructure for examining security properties of software. In *Proceedings of the 9th ACM conference on Computer and communications security* (CCS '02), Vijay Atluri (Ed.), pp.235-244, ACM, New York, NY, USA.

Devanbu, P.T. & Stubblebine, S. (2000) Software engineering for security: a roadmap. In *Proceedings of the Conference on The Future of Software Engineering* (ICSE '00). ACM, New York, NY, USA, pp.227-239.

Dhiman, K. et al. (2013) Clojure: modular programming with functional, concurrent language on the JVM. In *Proceedings of the 2013 Conference of the Center for Advanced Studies on Collaborative Research* (CASCON '13). IBM Corp., Riverton, NJ, USA, pp.370-371.

ENISA (2013) *ENISA Threat Landscape 2013*. [online] Available at: <http://www.enisa.europa.eu/activities/risk-management/evolving-threat-environment/enisa-threat-landscape-2013-overview-of-current-and-emerging-cyber-threats/at_download/fullReport> [Accessed 2.1.14]

GIT (2013) *Georgia Institute of Technology "Emerging Cyber Threats Report 2014"*. [online] Available at: <http://www.gtcybersecuritysummit.com/2014Report.pdf> [Accessed 13.1.2014]

Gordon, L.A. & Loeb, M.P. (2002) The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, 5(4), pp.438–457.

Groom, P.D. (2003) The IT security model. *Potentials*, IEEE, 22(4), pp.6-8.

Hintzbergen, J., Hintzbergen, K., Smulders, A. & Baars, H. (2010) *Foundations of Information Security: Based on ISO27001 & ISO27002.* Zaltbommel: Van Haren Publishing.

ISACA (2011) *CISM Review Questions, Answers and Explanations Manual 2012*. ISACA .

ISO (2001) *ISO/IEC 9126-1:2001 Software engineering -- Product quality -- Part 1: Quality model.* Geneva: ISO.

ISO (2011) *ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models*. Geneva: ISO.

Krauss, C. & Fusenig, V. (2013) Using Trusted Platform Modules for Location Assurance in Cloud Networking. In *Network and System Security*. Springer, pp.109–121.

Kushida, K.E., Murray, J., Zysman, J. (2011) Diffusing the cloud: Cloud computing and implications for public policy. *Journal of Industry, Competition and Trade*, 11(3), pp.209–237.

Marotta-Wurgler, F. (2007) What's in a Standard Form Contract? An Empirical Analysis of Software License Agreements. *Journal of Empirical Legal Studies*, 4(4), pp.677–713.

Marotta-Wurgler, F. (2011) Some realities of online contracting. *Supreme Court Economic Review*, 19(1), pp.11–23.

OWASP (2010) *OWASP Secure Coding Practices Quick Reference Guide*. [online] Available at: <https://www.owasp.org/images/0/08/OWASP_SCP_Quick_Reference_Guide_v2.pdf> [Accesed 19.1.2014]

Perera, G. (2012) *Purposefully manufactured vulnerabilities in US government technology microchips: risks and homeland security implications*. Ph.D. Naval Postgraduate School.

Pfleeger, C.P. (1997). The fundamentals of information security. *Software*, IEEE, 14(1), pp.15-16.

PwC (2013a) *PricewaterhouseCoopers: The Global State of Information Security® Survey 2014*. [online] Available at: <http://www.pwc.com/gx/en/consulting-services/information-security-survey/download.jhtml> [Accessed 2.1.14]

PwC (2013b) *PricewaterhouseCoopers: China's impact on the semiconductor industry 2013 update.* [online] Available at: <http://www.pwc.com/gx/en/technology/chinas-impact-on-semiconductor-industry/assets/china-semicon-2013.pdf> [Accessed 11.1.14]

Rainer, R.K., Marshall, T.E., Knapp, K.J. & Montgomery, G.H. (2007) Do Information Security Professionals and Business Managers View Information Security Issues Differently? *Information Systems Security*, 16(2), pp.100-108

Reuters (2013) *Exclusive: Secret contract tied NSA and security industry pioneer*. [online] Available at: <http://www.reuters.com/article/2013/12/20/us-usa-security-rsa-idUSBRE9BJ1C220131220> [Accessed 11.1.14]

Schroeder, N.J. & Grimaila, M.R. (2006) Revealing Prospect Theory Bias in Information Security Decision Making. In *Emerging Trends and Challenges in Information Technology Management: 2006 Information Resources Management Association International Conference, Washington, DC, USA, May 21-24, 2006*. pp.176-179.

Shostack, A. & Stewart, A (2008) *The New School of Information Security*. Boston: Addison-Wessley.

Skorobogatov, S. & Woods, C. (2012) Breakthrough silicon scanning discovers backdoor in military chip. In *Proceedings of the 14th international conference on Cryptographic Hardware and Embedded Systems*. pp.23–40.

SST (2012) *Solid State Technology: Top 12 semiconductor foundries of 2012*. [online] Available at: <http://electroiq.com/blog/2012/08/top-12-semiconductor-foundries-of-2012/> [Accessed 11.1.2014]

Tarnovsky, C. (2008) *Black Hat DC - February 21, 2008 Presentation: Security failures in secure devices.* [online] Available at: <http://www.blackhat.com/presentations/bh-dc-08/Tarnovsky/Presentation/bh-dc-08-tarnovsky.pdf> [Accessed 28.1.2014]

Tsiakis, T. & Stephanides, G. (2005) The economic approach of information security. *Computers & Security*, 24(2), pp.105–108.

The Opportunities of National Cyber Strategy and Social Media

Arto Hirvelä[1], Aki-Mauri Huhtinen[1] and Tommi Kangasmaa[2]
[1]National Defence University, Helsinki, Finland
[2]Defence Command, Helsinki, Finland
arto.hirvela@mil.fi
aki.huhtinen@mil.fi
tommi.kangasmaa@mil.fi

Abstract: Securing the society is a central task of the state. On the present day, as well as in the future, knowledge and information are ever more closely tied to electronic data transfer. Finland's newly published Cyber Security Strategy depicts how the government safeguards electronic data transfer, that is, information security against different threat and risk scenarios. Cyberspace has humane element and technological element. It is a way to influence and effect society. It may be used to influence minds or attack physical world for example by disrupting traffic control. But cyberspace cannot exist without people. Strategic communication is a concept that unites efforts of governmental organizations to influence people in support of national interests. Strategic Communication is in a heated discussion in the military field. How the militaries can be credible and uphold the high standards of democracy within the asymmetric and complex battlefield. To be credible one must act according to what is said. Challenge is that the scene is global in the information age. The act, the actor, the scene, the purpose – all are exposed in a very short notice to the global audience through cyberspace. The most effective way to show the scene and actors is an audiovisual product. Examples of this come from all conflict zones. Militaries are establishing Youtube channels and providing material from the intense fighting and frontline action. The solution has been the Combat Camera capability. Media trained soldiers who are in the place where main effort is taken. Within spirit of strategic communication, Cyber strategy itself is a one way to operationalize cyber security by announcing measures to be taken against cyber attacks. In this article we take phenomenological reflections on the Finnish Cyber Strategy and Strategic Communication.

Keywords: cyber strategy, cyber security, strategic communication, combat camera

1. Introduction

Cyber security means the desired end state in which the cyber domain is reliable and in which its functionality is ensured. As a concept "cyber" refers to steering or connecting. Thus it is an act, for instance, to transfer information from analogical to digital form or from electronic to non-electronic form. Finland's Cyber Security Strategy slightly over-emphasises the information security aspect of cyber. It tries to find technical solutions to technical problems, even if psychology would be a better remedy.

The Internet is a central dimension from a cyber security perspective. The operation mode of the Internet world has to a large extent evolved outside state steering. Therefore, governments are experiencing diminishing opportunities of technological and legal control in relation to the expansion of the Internet world. The copyright fee paid when buying a CD is an example of the codification of technology neutral norms that have psychologically dubious consequences.

Exercising political control over the Internet is difficult, because the Internet world lives off vertical and horizontal interdependence. Even minor restrictions to people's freedom of action generate quick counter reactions that are poisonous to political activity. Hacktivism may also turn against itself. The Snowden case points out that even if defending truth and freedom is a good aim in itself, it often generates unwelcome counter reactions, such as erosion of trust. When secrets become revealed, nobody can truly trust anyone. In information networks truth turns into gossiping as it becomes multiplied and its message waters down.

According to German Army Manual for Unit Command in World War II "Das FuG" (Condell, Bruce & Zabecki 2009), war is an art, a free and creative activity founded on the scientific principles. It makes the highest demands on human personality. When we study the history of art of war we notice that the character counts more in a war than intellect. Often incalculable elements have a decisive influence to the victory or defeat. For that reason we have to first understand the binaries in Battle. The armies and security forces are always reflections of the societies from which they are drawn. This important statement is also connected to our information age. Cyber domain is also something new in the warfare and generally in security activities and practices. Some things in warfare are measurable and calculable but still more can only be assessed qualitatively.

2. Cyber security

Cyber attacks can seriously disrupt or even paralyse segments of critical infrastructure and society's vital functions. According to Finnish Government Cyber Security concept "Cyber security is not meant to be a legal concept the adoption of which would lead to granting new competences to authorities or other official bodies. In this respect, no changes are proposed to the basis of contingency arrangements or to regulations concerning the competences of authorities." (FCSS 2013) Thus far cyber operations have been interpreted as 'soft measures,' for which reason the threshold for using them is estimated to be below that of traditional military operations. The increasing cyber activism, cybercrime and cyber espionage denote growing activity among states and non-state actors. Consequently, the cyber domain has transformed the traditional power structure, providing even small states and non-state actors with an opportunity to have effectual action. In cyberspace, it is no longer size and mass that matter, rather it is expertise.

Afore described principle forces us to rethink the relationship of management and administration to the cyber world. The concept of cyber remains outside juridical and normative procedures, which indicates a necessity to change them. For instance, authorities need to plan everything in such a way that it tolerates even the negative interpretations that publicity brings forth. How can a task that is not covered by the official norms be carried out as one of the state's administrative tasks? Security of the cyber world cannot be increased by creating more norms. The administrative style and attitude of the Strategy are not the correct response to a need for cyber collaboration. The more cyber action becomes restricted and denied, the more these restrictions and prohibitions are circumvented.

The vision of the Cyber Security Strategy is mostly based on collaboration between the Finnish authorities and on the cultural spirit of consensus. Nevertheless, in the Internet era cooperation between the authorities has not been a central driver of networking within civil society. Instead of governmental normativity the authorities need new tools for leadership and management, such as the Comprehensive Approach. The Approach is based on different principles than the assumption that each administrative sector is responsible for its own cyber security and that the government takes care of the entirety. What might be needed is the so-called model procedure of strategic communication.

It may even be claimed that cyber threats emerge from the inability of legislation to control civil society any longer. It is more relevant to accept the fact that cyber threats reside within the functions vital to society in the form of potentiality than to try addressing them by the means of traditional risk analysis. When the potentiality realises itself, it is essential to return the functions to their normal operating mode or to redirect them to alternative channels. It would be necessary to move from "if problems emerge" way of thinking to the "when problems emerge" way of thinking. The difference in psychological nuance is remarkable.

Technology always contains the seeds of its own destruction. Cyber resilience is a relevant concept. What is essential in it, however, are not the traditional preparations done beforehand but the transfer of resources to recovery when the threat realises itself. Then, for example, cyber resources are not in the possession of the public authorities, but reserved for organisations like the social insurance institution of Finland to help civil society. The question is whether cyber resources eventually benefit civil society or are they used to secure some minor, elitist "vital" functions. Who acknowledges the realisation of cyber threat and directs the overall resources towards recovery? Whose cyber security does the strategy deal with and with what kind of means? Who is allowed to use cyber security power?

To a large extent cyber security is the next phase in the continuum after information security. However, cyber becomes reality more often in a non-technological form, for instance, as in the functions of social media. Previously, computer was visible while action was invisible, but today technology and machines are hidden behind functionality. Thus, cyber security requires ever more know-how which comes from outside the technical sector and education.

In social media, it is no longer the facts that matter, but the meaning these facts are given to in human interplay. A large technological problem may cause little psychological frustration. With regard to cyber security it is also necessary to ask, whose fears are worth relieving. Information is not transferred to those who need it (thus, uncertainty is created within the population who does not know what has happened) or it is manipulated. Majority of information on the Internet travels around without any kind of integrity or

provenance check. Thus it is relatively easy to manipulate the information received even by the masses without anyone noticing it. People's opinions become polarised and regroup around extreme attitudes, while "normal" opinions disappear.

The organisational programmes of decision (rules, regulations, procedures and code of ethics) work as safety-nets. This means that the safety-nets protect organisations from the emergence of moral questioning and thereby shield actors from the burden of personal responsibility. As we can never know if we have discharged our responsibility, we must always doubt that our actions are ethical. The only way of the manifestation of truth is a readiness to speak in one's own name and take a risk by speaking out against practices which are perceived as unjustifiable. (Weiskopf 2013)

3. Military cyber defence

The fifth strategic guideline in the definition of policy includes the basic principle of how the Defence Forces establish cyber defence capability as part of their statutory tasks. The aim is to secure their own command and control systems so that the statutory tasks can be carried out irrespective of cyber threats. In addition, cyber related intelligence and the capability to impact will be developed as part of the overall military capability. Legislation and cyber defence training will be practiced as parts of the activity.

Cultural change and especially the linguistic turn are central also in the military discourse. This generates pressures for the transformation of procedures and leadership within the Defence Forces. It is expected to become more social and attend public networks in order to bring its message forth and thus, to pre-empt. (Rantapelkonen & Salminen 2013)

Western military forces are able to protect and secure their command and control systems, but strategic partners and subcontractors who carry out several technological and maintenance tasks pose a challenge. The latter do not operate on the basis of security production per se, but pursue economic profit. For example, the new flagship of fighter defence, the extremely expensive and extremely technological F-35 fighter, is built upon the guidance of millions of computers and constant software updates. How can it be secured that software does not contain back doors and that some kind of collapsing domino effect will not occur? The fighter is piloted in a way an iPad is used. It is not necessary that the potential offender fully understands the target system. It is enough that one understands how the system can be disrupted.

The Defence Forces are planning and preparing cyber defence. Nonetheless, there is more than one way of conducting cyber war. The cyber domain is embedded in different ways of warfare. Cyber warfare breaks the traditional image of a united and hierarchical military organisation, because cyber war is a mixture of old and new, as well as of actors operating on different levels. It is possible to begin a cyber war without the challenging preparation time required by traditional weapon systems.

In the future, the Nordic countries will be ever more dependent on technology, electricity and computers, as well as on networks. Therefore, cyber warfare is part of their way of war. States and their networks are more likely to develop offensive cyber war doctrines and means than individuals are. On the other hand, several treaties restrict the actions of states, but do not bind non-state actors. Stuxnet triggered states to develop offensive capabilities. The leading states are advancing modular cyber weapons, constantly updating their software, and sharing different components amongst each other. A reason for hiding cyber warfare capabilities may be the lack of offensive capabilities or their incomplete state.

Risks related to the digital planning and materialising of different substances pose a great challenge for the Defence Forces. It is exactly malware like Stuxnet that may influence the strength of missiles and different weapons material, and hence turn them dangerous to use. The cyber domain has also increased the usability of nuclear weapons. A dirty bomb made out of nuclear weapons can disable the entire cyber domain in time and place by using the electro-magnetic pulse.

Commercial companies have a far-reaching opportunity to arm cyber war and sell codes, for instance, for the elaboration of bot-networks, such as the Russian Business Network. The best way to shelter is to develop a command and control system that protects all processes and applications in use. The only way of defending is cooperation between different actors and compromises between the procedures of different actors.

Even if the state is cyber defensive at the political level, the generation of offensive capabilities should be allowed at the practical level. This is necessary for the achievement of the strategic goal. In order to shelter from a cyber attack one needs to consider how an attack would be conducted. In cyber battle offense and defence merge into one functional entity.

4. Cyber defence in Finland

According to FCSS (2013), Finland is one of the most developed information societies whose function relies on various electronic networks and services. Finland has already been the target of cyber operations where the focus was on cyber activism, cybercrime and cyber espionage. The international development in cyberspace increases the possibilities of new threats being used against us. The public administration and the business community are continually being targeted by crackers and hackers attempting to exploit system vulnerabilities. That the targets are carefully selected and studied only serves as an indication of the professionalism of the attacks. Sophisticated malware and techniques are increasingly used in these attacks.

The military dimension of cyber security, that is, cyber defence is mentioned in the Strategy. It is aligned that "[t]he Finnish Defence Forces will create a comprehensive cyber defence capability for their statutory tasks." Military cyber defence encompasses intelligence, as well as influence and protection capabilities.

The aim of cyber intelligence is to provide pre-warning and to facilitate targeting. From the perspective of cyber defence it is necessary to collect information on weapon systems and networks used by a hostile actor, as well as on their vulnerabilities. Cyber influence capability entails the capability to conduct cyber operations on the systems and networks of the hostile actor as part of overall military operations.

The capability to shelter refers to the protection of the activities of the Defence Forces against the intelligence and influence operations of the hostile actor. This requires around-the-clock maintenance of situational awareness and monitoring, collaboration with other authorities and management of the organisation's information security. For example, the phone numbers currently in use reveal directly the hierarchical structure of the organisation.

The Defence Forces have improved their capabilities to create situational awareness and shelter in the information dimension which has followed the societal development of information technology. The utilisation of capabilities calls for jurisdiction. There is a need for political decision making on the rules of engagement with regard to cyber influence. We may ask, what are the international decrees on the rules of cyber engagement? How cyber force is identified?

Improving cyber capabilities entails additional resources and, especially, the enhancement of people's skills and know-how. Material expenses are primarily based on the continuous improvement of training environment, the acquirement and maintenance of tools for creating situational awareness and analysis, as well as the development of technical procedures and tools. However, it cannot be taken for granted that in Finland there are enough people learning the necessary skills to conduct effective cyber defence.

Nonetheless, jurisdiction is the most urgent issue in cyber defence. It should be scrutinised in the international context in which other European states and their armed forces develop the aforementioned capabilities.

5. Strategic communication in cyberspace

Cyberspace has humane element and technological element. It is a way to influence and effect society. It may be used to influence minds or attack physical world for example by disrupting traffic control. But cyberspace cannot exist without people.

Strategic communication is a concept that unites efforts of governmental organisations to influence intended key audiences in support of national interests. The concept tries to get actors and agencies at all levels to work towards strategic level ends by synchronising messages. It tries to connect actors at all levels both horizontally and vertically.

Sirén (2013) understands strategic communication as a proactive and mind-centric paradigm of the art of war and as a maxim of the paradigmatic evolution of the art of war. Strategic communication is therefore a

reflective effort to construct a worldwide potential of tolerance into human communities and societies in order to avoid future wars and armed clashes without losing peace. This is based on a theory of positive recognition and done by shoving example of what being wonderful and acting accordingly means. (Sirén 2013)
Cyberspace is one of the global commons. With increased electronic social cooperation and interactivity, the ability to communicate with most of the world's population has become easier. Free access to global information through information systems, such as traditional media and Internet, is vital for fulfilling the human need for self-actualisation. Free access to global information is also a way for tolerant nations to build a truthful and tempting model of society and communicate that to the target audiences, and also challenge repressive governments, which try to prevent the use of it in many ways (Sirén 2013).

Communities need to establish their credibility in information environment. Modern selective receivers of information need more assurance that the source of any information is credible and trustworthy. Obtaining the endorsement of an objective third-party gives senders the credibility to be better believed by targeted audiences. Still, communities must work hard, so that people would be able to understand each other. Even when operating in cyberspace the most important interaction is still the one between people, not the one between computers (Seppänen 1998).

6. Social cyberspace and combat camera

Visuality has increasing importance in strategic communications. Access to operations is easier when the mental battleground has been prepared thoroughly. The language of the picture is global. It is *lingua franca* and thusly very convenient for the military which operates globally. Even the defending armies are in the need of influencing and informing the global audience in the time of crisis. The challenge is that social media and Internet provides hundreds of solutions for disseminating video and pictures all over the conflict area. Whose view gets first to the Internet? Who is the leading voice and producer?

Combat Camera capability has been the answer for the battle between the combatants in the audiovisual battlefield. They have the same basic capabilities as all civilians and combatants with their cellular phones and cheap video cameras. The difference lies elsewhere. First of all, the combat camera soldiers are media professionals. The combat camera soldiers are hired from the ranks of young media trained students and professionals. They are trained by information operators and communication specialist to provide the messages that are important to the nation's narrative. They also have means to provide material to those who need it even when the normal communication lines are down. These soldiers are also trained to go to the tightest spot to get the material. Even with the navy, air force or special forces.

When nation starts to fight on the mental front, it has to utilize all the resources and channels it can get in the Internet domain. This domain is full of visual information, right from the start of the conflict. Good examples have been the war between Israel and Lebanon in 2006 or Gaza conflict in 2012. It has been true for the war in Syria from the start. These conflicts show us that there is an ongoing process to utilize combat camera capabilities to fill Youtube and other video services with the own message. The militaries have noticed that fragmented messages, taken with the phone can't convey the controlled and planned effect. They are like cries from the area and hard to steer into lasting effect. The message can't be different each time. This fragmented approach can backfire easily and is very much non-military like way to approach battle space. The answer has been more controlled production, which is integral part of the communications plan. It fills the information operations need for tools to convey the message.

7. Militarization of social media

There is no ethical norm that could sweep military outside of the Internet and its most important mind twisting tool – the social media. This means that the messages coming from the conflict areas are more and more organised and framed. The more you get the professional capabilities like the combat camera, the more these messages will fill the cyber domain (McKeldin 2009). Building of these messages will follow the rules and inventions of the rhetoric and framing. Theoreticians like Goffman or Perelman have shown their importance in building the combat camera thinking. The power lies in influencing by showing the truth among the many.

The social media is very vulnerable to influencing. The messages are disseminated efficiently and quickly by sharing functions. The more visually attractive, the easier and quicker is the message influencing the global mindset. The strength lies in the concept of truth. The picture "can't lie." It is prone to manipulation of course,

but this method is also prone to thousands of judges who will scrutinize the material and expose this to the world. The producer loses the credibility and the following products are not so effective any more. Those who disseminate "truths" that are credible, visually attractive and include the big story in small bites will win the hearts and minds of the audience.

Combat camera productions are truths among the other hundreds of truths. They are easily defended as direct manipulation is totally unacceptable in the rules for Combat Camera. The picture, being true is the result of framing, storytelling, timing, perception, professionalism and influencing. The final product is as much "true" as it can be but the same time it is as "false" as it is wanted (Sirén 2013). The same is true for journalist in the conflict zone. They follow probably high ethical codes of those in journalist profession but at the same time they face the demands of the economy and the challenges of the crisis zone (Andersen 2006). The truth they provide is affected by their subjective vision of the conflict and framed by their own view of the situation. This dilemma is shown from time to time as the most effective pictures from the conflict areas are shown in larger frame. These reveal the larger set-up, which is often very different from the feeling the picture has shown us. The same phenomenon is sometimes visible in combat camera productions.

The difference between those who just take the picture and send it to the Internet and the combat camera lies in the process of framing the truth. Seasoned combat camera producers and operators are very good at telling the big story. Their support for the nation's narrative is critical. They have the capability to get what is wanted from the most dangerous "stages." Their ability to visually tell a story is based on vigorous training and media professionalism. They are capable of delivering the material quickly and in the right format. They have means to disseminate the products through channels that have been already used during the peace time. Those channels are already credible and tested. The story is consistent and looks right. The products tell the same message, which is part of the bigger document called the Strategic Communication Plan.

8. Conclusion

Cyber politics may be guided nationally, but technology or legislation does not yield to similar control. The two latter ones are global forums. The principles of the operation mode presented in the strategy include a challenge to the management of the social. In the principles it is noted that "[c]yber security arrangements follow the division of duties between the authorities, businesses and organisations, in accordance with statutes and agreed cooperation". This view of leadership is a view of administration, but hardly that of civil society which lives and functions off the Internet. In the principles the technological and functional starting points for the development of cyber security are also emphasised, albeit the focus should be on the social development and coordination of security.

In cyber world, size and mass no longer dominate, but know-how does. The great powers possess significant resources for operating in cyber operational environment but they do not dominate it. Small actors may acquire sufficient skills and resources to take advantage of the opportunities residing in cyber operational environment.

Cyber security education cannot be left outside university degree programmes or be included only in ICT studies. Social and humanistic skills and know-how are also required. Cyber security know-how needs to be developed on all sectors of education. The general aim of cyber security training should be to increase the awareness of citizens, authorities and business of threats and risks that reside within cyber operational environment. In addition, the general skills required for an effective implementation of cyber security procedures should be enhanced. Investments should be made in the training of Internet know-how from the early childhood onwards.

Cyberspace is one of the global commons and in the modern age vital for fulfilling the human need for self-actualisation. Cyberspace enables nations to communicate a truthful and tempting model of society to the target audiences. There is a battle for narratives fought in cyberspace.

The establishment of the combat camera capability is part of the inevitable process of militarisation of the social media. The social media is so powerful weapon in the battle for the human mind that it can't be left outside of the military capabilities and functions. The battle between two sides is always about enforcing the other to accept your values, morals and goals.

The main challenge with social media is the fast spread of information without restrictions. Once the information is in cyberspace it is impossible to control its distribution. Strategic Communication needs to assure that presented narratives are truthful and consistent.

References

Andersen, Robin (2006). A Century of Media, *A Century of War*. Peter Lang Publishing. p 317

David, G.J. & McKeldin, T.R (2009) Eds. *Ideas as Weapons*. Potomac Books, Washington D.C. p 135

Condell, Bruce & Zabecki, David T (2009) Eds. *On the German Art of War*. Truppenführung. German Army Manual for Unit Command in World War II. Mechanicsburg PA, Stackpole Books. p 18

Finland's Cyber Security Strategy (FCSS) (2013). Background dossier. Available at www.yhteiskuntaturvallisuus.fi

Israel Defence Forces Youtube-channel. www.youtube.com/idfnadesk. Accessed 12.2.2014.

Rantapelkonen, Jari; Salminen, Mirva (2013). *The Fog of Cyber Defence*. Rantapelkonen, Jari; Salminen, Mirva (Eds). National Defence University, Department of Leadership and Military Pedagogy, Series 2: Article collection N:o 10, Juvenes Print Oy, Tampere

Seppänen, R. (1998), "Johdanto" in *Tietoyhteiskunnan Harha (Illusion of Information Society)*, Seppänen, Rauno (Ed), Kustannusosakeyhtiö Puijo, Kuopio, p 11

Sirén, Torsti (2013). *Winning Wars Before They Emerge - From Kinetic Warfare to Strategic Communications as a Proactive and Mind-Centric Paradigm of the Art of War*, Universal-Publishers, Boca Raton, pp 6, 8, 119, 258

Weiskopf, R; Willmott, H. (2013). "Ethics as Critical Practice: The "Pentagon Papers", Deciding Responsibility, Truth-telling, and the Unsettling of Organizational Morality". *Organization Studies* 34(4), 469-493.

Online Social Networks: A Vehicle for Malware Propagation

Ehinome Ikhalia[1] and Johnnes Arreymbi[2]
[1]Sleeky.net, London, UK
[2]School of Architecture, Computing and Engineering, University of East London, London, UK
Ehinomeikhalia@sleeky.net
Jarreymbi@gmail.com

Abstract: Online users have completely changed the initial concept of social networks and how they have been perceived in recent years. Online Social Networks (OSNs) have completely changed in form, scope and nature, by the way Internet users create, view, share, present and/or use information. As a result, many vulnerabilities have emerged and the security threats, attacks, and malware of OSNs continue to increase because malicious or malevolent users commonly referred to, as 'attackers' often take advantage of the inferred trust relationships inherent in social networking to promote cybercrimes. This paper aims to investigate commonly used approaches to malware propagation on online social network systems, by examining their attack vectors and present a qualitative review of the findings. The study shows that online social networks are still vulnerable to Cross site scripting (XSS) and Clickjacking attacks. We further show that Clickjacking using malware URL links is still the dominant source of malware propagation on online social networks. Finally, we propose areas for further research used to stem the use of OSNs as a medium for malware distribution.

Keywords: malware, cybercrime, online networked security, clickjacking, drive-by-downloads, social media, cross site scripting, Sleeky.net, Facebook, Twitter

1. Introduction

Online social networks (OSNs) are growing exponentially as it remains the most popular communication infrastructure for Internet users (Ikhalia, 2013). Simultaneously, malware that specifically target these online social networks are also escalating. Online social networking has totally reinvented how Internet users create, view, and share information as millions of users now have kind of cyber presence. As a result, the security threats, attacks, and malware of OSNs will continue to increase as malicious users take advantage of the inferred trust relationships inherent in social networking. Platforms providers, such as Facebook and Twitter, are now the biggest targets for malicious users, spammers, and other cybercriminals looking for easy victims, making online social networking the largest distribution platform for malware (Leitch & Warren, 2009).

Facebook now has over one billion active users, making it very eye-catching for cybercriminals to reach a large number of possible victims (Ikhalia & Imafidon, 2013). Moreover, the relationship model of online social networking permits the rapid distribution of malware; which could be installed on a victim's computer and trick victims into revealing sensitive data (e.g., credit card information, bank information, and other personal data used for identity theft) Arreymbi, (2005a), (2005b); or make users complete unnecessary surveys which earn the cybercriminal affiliate promotion income (Thomas & Nicol, 2010).

2. Related work

Previous researches have investigated the propagation of malware through online social networks. Yan, Chen, Eidenbenz, & Li, N (2011) investigated the characteristics of malware propagation in online social networks based on datasets collected from a real-world location-based online social network. They analysed the social structure and user activity patterns of this network, and agreed that it is a typical online social network. It suggests that conclusions drawn from this specific network can be translated to other online social networks. They also used extensive trace-driven simulation to study the impact of initial infection, user-click probability, social structure, and activity patterns on malware propagation in online social networks. The goal of their work sheds light on the characteristics of malware propagation in realistic online social networks. However, the approach was too generalised. Their experiments underestimated malware propagation speed. The user-click probability in their experiments was the same over all the users. This is unrealistic, given the fact that different online social network users have different awareness levels on malicious embedded URLs sent from their friends.

Thomas & Nicol (2010) explored Koobface's zombie infrastructure and analysed one month of the botnet's activity within both Facebook and Twitter. Constructing a zombie emulator, enabled them to infiltrate the

Koobface botnet to discover the identities of fraudulent and compromised social network accounts used to distribute malicious links to over 213,000 social network users, generating over 157,000 clicks. They recommended that to stem the threat of Koobface and the rise of social malware, social networks must advance their defences beyond blacklists and actively search for Koobface content, potentially using infiltration as a means of early detection. The authors recommended that to stem the threat of Koobface and the rise of social malware, social networks must advance their defences beyond blacklists and actively search for Koobface content, potentially using infiltration as a means of early detection. Their work stressed generally that obfuscation will continue to pose a threat to social network defences, but not a specific technical solution was recommended to prevent users from clicking malicious URLs in social networks.

3. Cases of malware attacks on social network sites

According to a report by Securelist (2010), malicious users have exploited Twitter as a novel platform to perform their malevolent behaviour which includes distributing spam and phishing scams, spreading malware, hosting botnet command & control channels as well as launching other subversive criminal activities. In March 2010, cyber-criminals exploited Twitter to distribute malicious programs using festive-themed mails. Furthermore, in September later that year, thousands of Twitter users plus the wife of a former British Prime Minister and White House Press Secretary were compromised by Twitter's malicious users.

In May 2010 Facebook experienced a malware attack based on the 'Like' feature. The feature generates a list of things that an owner of a Facebook account 'likes'. Thousands of users became victims to the attack which was nicknamed 'Likejacking', a play on the term 'Clickjacking'. The attack involved posting a tempting link on Facebook, e.g., 'Watch The World Cup 2010 in HD' or '101 Hottest Women in the World' (Securelist, 2010). The link directed to a specifically made page, where JavaScript code placed an invisible 'Like' button anywhere the cursor happened to be. The button continued following the cursor and recorded each press irrespective of where the user really clicked. Consequently, a replica of the link was posted to the user's 'Wall' and a message which displayed that 'the user liked the link' displayed on the 'newsfeed' of other users connected as 'friends'. The malware attack had an escalating effect because the malicious link was followed by friends, then friends of friends and so onward. As soon as the link was added to the Wall, the user was readdressed to a page which contained the information promoted, such as pictures of girls. The page also hosted a slight piece of JavaScript code for a marketing campaign for the attackers who had set up the deceitful system with a small sum every time a user was redirected to the page.

4. Conceptual outlook of malware propagation pattern

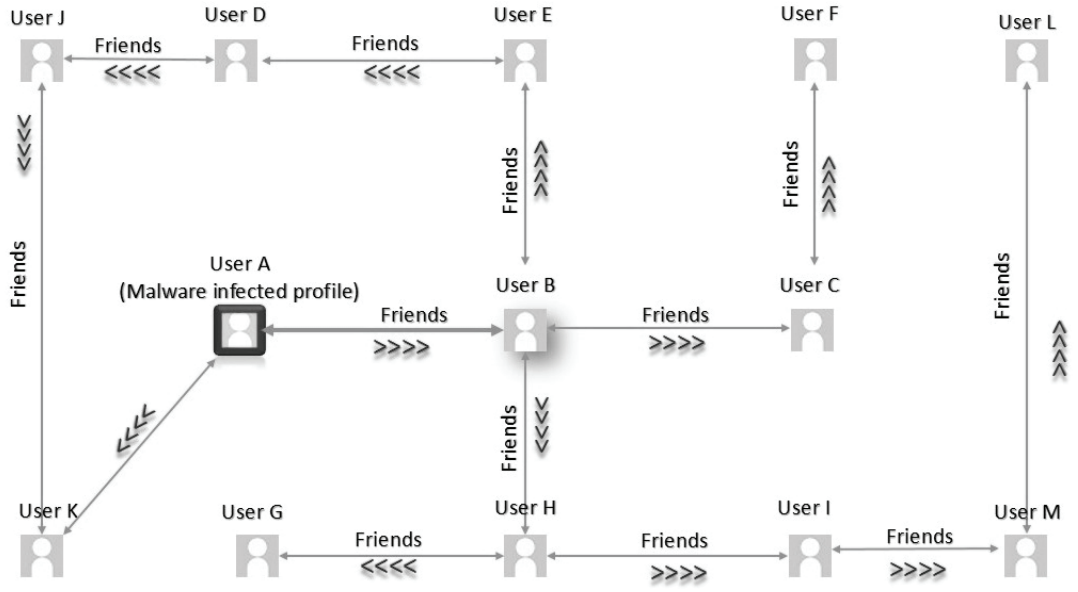

Figure 1: Malware proliferation patterns on online social networks

The open connection model of online social networks allows malware to spread rapidly and exponentially (Stein et al, 2011). Social networks characteristically display a great clustering feature; this means that if user A knows B and K, then B and k are most likely to know each other. Figure 1 provides an impression of the how

malware proliferates through vulnerable users on online social networks. The default avatar represents users within the system having interrelations. It also shows the pattern of interconnectivity and the chain infection progression on a typical social networking site.

Consequently, an infected **User A** can infect all the other interconnected users within the social networking environment (Yan et al, 2011). For instance, a malicious user profile in an online social network can infect the other user profiles that share a mutual connection as shown in figure 1..

5. Malware proliferation methods

5.1 Clickjacking

A Clickjacking attack is executed when a user is misled into naively triggering an invisible malicious script in a system while appearing to click a visible image link, video link or text link from a completely different system. Though being logged into some target system, the user visits the attacker's malicious site which displays the links that the victim interacts with. In actuality, the clickjacked page is a transparent layer above the visible image links, video links or text links with action controls that the attacker desires the victim to execute (Lundeen, 2012).

When the potential victim clicks on the image, video or text links on the page, it essentially activates the action controls in the transparent covering layer. Dependent on the type of action control, the malicious user deceives the victim into effecting a probably 'advantaged' functionality on the target system to which the victim is authenticated (Thomas & Nicol, 2010). The main issue is that there is always a dichotomy between what the victim assumes to have clicked against what was actually clicked (Mahmood, 2012).

Within a social networking environment, Clickjacking is a technique which tricks gullible users (victims) into clicking on a button or link, a good example is Facebook's 'like-jacking'. In this situation, users will be shown a video player that is alike to YouTube video. When the video is clicked, the video does not play, instead, the Facebook 'like' button of the content is being activated. Therefore, users are misled to 'like' the page in order to make it more popular. Furthermore, some of these fake video links may prompt users to input some sensitive data before viewing the video, so that the victim's information can be obtained by the attacker (Lundeen, 2012).

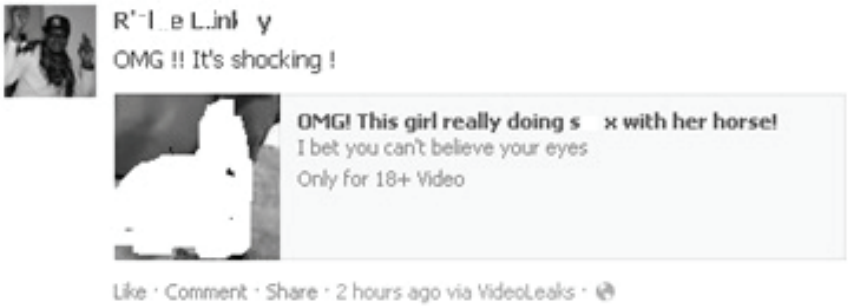
OMG !! It's shocking !
OMG! This girl really doing s x with her horse!
I bet you can't believe your eyes
Only for 18+ Video
Like · Comment · Share · 2 hours ago via VideoLeaks ·

Figure 2: An example of a malware video URL link on Facebook targeted at clickjacking an innocuous user (Source: Facebook, 2013)

Table 1 shows that clickjacking is the most used and active attack technique employed by malicious users in 2012. It contributed to 87.36% of the most active malicious programs involving Web attacks. And,the fact that the process exploits URLs, malicious users or attackers have now adopted this technique on vulnerable online social network users (Lundeen & Alves-Foss, 2012).

In order to simplify clickjacking attacks using figure 3, let A denote a legitimate Web page (e.g., social networking site) with a harmless user interface component A1. A malicious user develops a malicious Web page B with a component B1 to execute a malevolent action (e.g., making the victim to download malware unwillingly) by transparently overlaying a somewhat advantaged functionality B1 on top of A1. Therefore, when a potential victim clicks on A1 on the Web page A, the victim actually clicks on B1 on the Web page B, hence initiating an unintentional action.

Table 1: Top 20 most active malicious programs involved in web attacks on users' computers. (Securelist, 2012).

Rank	Name*	Number of attacks	% of all attacks**
1	Malicious URL	1 393 829 795	87.36%
2	Trojan.Script.Iframer	58 279 262	3.65%
3	Trojan.Script.Generic	38 948 140	2.44%
4	Trojan.Win32.Generic	5 670 627	0.36%
5	Trojan-Downloader.Script.Generic	4 695 210	0.29%
6	Exploit.Script.Blocker	4 557 284	0.29%
7	Trojan.JS.Popupper.aw	3 355 605	0.21%
8	Exploit.Script.Generic	2 943 410	0.18%
9	Trojan-Downloader.SWF.Voleydaytor.h	2 573 072	0.16%
10	AdWare.Win32.IBryte.x	1 623 246	0.10%
11	Trojan-Downloader.Win32.Generic	1 611 565	0.10%
12	AdWare.Win32.ScreenSaver.e	1 381 242	0.09%
13	Trojan-Downloader.JS.Iframe.cxk	1 376 898	0.09%
14	Trojan-Downloader.JS.Iframe.cyq	1 079 163	0.07%
15	Trojan-Downloader.JS.Expack.sn	1 071 626	0.07%
16	AdWare.Win32.ScreenSaver.i	1 069 954	0.07%
17	Trojan-Downloader.JS.JScript.ag	1 044 147	0.07%
18	Trojan-Downloader.JS.Agent.gmf	1 040 738	0.07%
19	Trojan-Downloader.JS.Agent.gqu	983 899	0.06%

5.1.1 The process of clickjacking

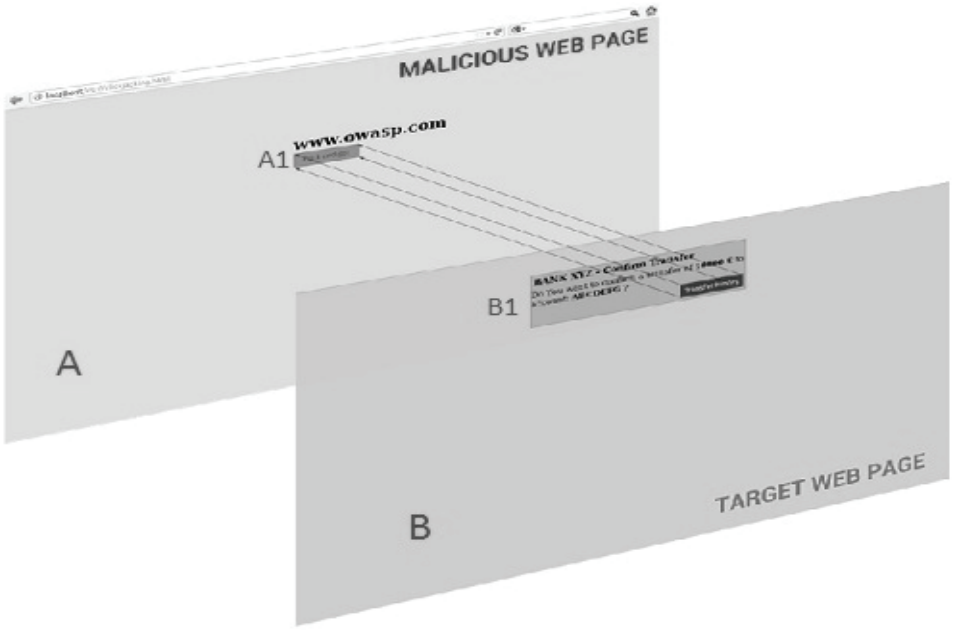

Figure 3: An example of a clickjacking attack

Clickjacking attacks can also executed using more sophisticated techniques as follows:

- By utilizing the Cascading Style Sheets (CSS) cursor property, malicious users can hide the default cursor and technically draw a counterfeit cursor somewhere else (Kim & Kim, 2013), or otherwise set a conventional mouse cursor icon to a misleading image that has a cursor icon shifted quite a few pixels off the original position.
- Malicious users can also use JavaScript (a single click can be transformed into a double click which can click on **A1** as well as **B1** concurrently).
- Clickjacking can be utilized with additional attack such as Cross-Site Request Forgery (CSRF), an extensively exploited Website susceptibility whereby illegal commands are communicated from a user that is trusted by the Website (Mohtasebi & Dehghantanha, 2011).

5.1.2 Cross site scripting

Cross-site scripting (XSS) can be described as an attack that takes advantage of susceptibilities in the codes of a Web application which allows a malicious user send malicious contents from an end-user and steal sensitive data from the victim (Leitch & Warren, 2009).

There are two types of Cross-site scripting attack, (1) persistent and (2) non-persistent (Ter Louw et al, 2009). Persistent or stored attacks involves the process whereby the injected malicious code is stored on the targeted server as HTML texts in a database system, a comment field or private messages sent on social networks. A vulnerable user's computer then accesses the malicious code on the server when it retrieves the stored information through a Web browser. Non-persistent or reflective attacks includes the process in which the injected malicious code is sent back to a user by the server in an error message, a search result, or any other kind of reply that reflects some of the user's input in that particular search result (Garg & Singh, 2013).

A Web page consists of texts and HTML mark-ups created by the server and construed by the client browser (Kim & Kim, 2013). Static Web pages have total control on how the Web browser reads these pages; in contrast, dynamic Web pages do not have control of how they are construed by the client. The core issue is that, if a malicious content is inserted into a dynamic web page, neither the website nor the user has sufficient information to recognize the occurrence in order to take defensive actions (Huang & Moshchuk, 2012).

Cross Site Scripting allows an attacker to insert malicious JavaScript or HTML code into a susceptible dynamic Web page to trick the user, implementing the script on the victim's computing device in order to steal sensitive data (Garg & Singh, 2013). The usage of Cross Site Scripting compromises private information, manipulate browser cookies, and makes requests which can be misguided for those of a legal user (He et al, 2011) (Neumann, Barnickel & Meyer, 2010).

Most Cross site scripting manipulation comprises sending the cookie to the malicious user through JavaScript codes, which dynamically sends the cookie to a Web server under the control of the malicious user (Xu et al, 2010; Mahmood & Desmedt, 2012). Figure 3.2.2 below shows a good example of a malicious JavaScript code that can be used to execute a Cross site scripting attack.

```
<script>
document.write ("<img src='http://attackersite/a.gif?x=" + escape(document.cookie + "'>")
</script>
```

Figure 4: A JavaScript code used for cross site scripting attacks

5.1.3 Attack paths of cross site scripting on social networks

Online social networks are extremely dynamic and packed with ever user generated varying content. The millions of users of social networks often pull data from many different sources. This data is incorporated with the platform and most times contains simple texts, images, videos and URL links (Thomas et al, 2012).

Often times the attackers uses the 'input fields' of the site to insert a 'comment' or 'post' that contains a malicious script (Huang et al, 2012). Any user who views that comment or post will automatically download the malicious script which is executed on the victim's browser, producing unwanted actions (Faghani & Saidi, 2009). A Facebook post on a user's wall or profile could contain a malevolent script, which can be executed on the browser of any user (especially the victim's 'friends') who visits the infected Facebook profile as illustrated in figure 3.1.

6. Conclusion and further research

As Online social networking continues to mutate, and changed the way Internet users create, view and share information, malicious users are also now using this ever sophisticated platform as a vehicle for malware proliferation. Our research explored the operations of Clickjacking and cross site scripting (XSS), which are two popular methods of malware proliferation, a technique largely adopted by malicious users. From this qualitative review, we conclude that establishments need a security technology that can analyse malicious URL links, since the link path is new and lacks identifiable signature or well-known payload. New but popular and ever challenging technologies such as online social networks, requires real-time content filtering security, which analyses information on the go, as it is generated and consumed. For example, envisage that a link is posted to a social networking site such as Facebook or Twitter and it directs users to a website that downloads or leads to data-stealing code through obfuscated JavaScript. This massive vulnerability can be controlled if further research is directed towards developing a security application to filter URL links posted by users on social networking sites.

References

Arreymbi, J. (2005a), *Online Banking: Spoofing Scams exposes Security Loopholes*. In Securing Electronic Business Processes. Paulus, S., Pohlmann, N., and Reimer, H.; (Eds). ISSE 2005, Vieweg publishers. Germany 2005. ISBN:3-8348-0011-2

Arreymbi, J. (2005b), *Phishing Attacks – A Socially Engineered Threat to e-Business*. Proceedings of the International Conference on Internet Computing. Las Vegas, Nevada, 2005

Faghani, M. R., Saidi, H. (2009). "Malware propagation in online social networks". *In Proceedings of the 4th IEEE International Conference on Malicious and Unwanted Software, Montreal, Canada, October*

Garg A., Singh, S. (2013). "A Review on Web Application Security Vulnerabilities". *International Journal of Advanced Research in Computer Science and Software Engineering. 3 (1), pp222-226.*

He, M., Horng, S. J., Fan, P., Khan, M. K., Run, R. S., Lai, J. L., & Sutanto, A. (2011). "An efficient phishing webpage detector". *Expert Systems with Applications*, *38*(10), 12018-12027.

Huang, L. S., Moshchuk, A., Wang, H. J., Schechter, S., & Jackson, C. (2012). "Clickjacking: attacks and defences". *In USENIX Security Symposium. USENIX Association.*

Ikhalia, E., Imafidon, C.O. (2013). "The need for two factor authentication in social media". *Proceedings of the International Conference on Future Trends in Computing and Communication -FTCC 2013. July 13th-14th, pp76-82.*

Ikhalia, E.J. (2013). "A New Social Media Security Model (SMSM)*". International Journal of Emerging Technology and Advanced Engineering. 3 (7), pp16-21.*

Imafidon, C.O, Ikhalia, E. (2013). "The investigation and implementation of social media security". *Proceedings of the 2nd global conference, London on communication information science and engineering. 24th to 26th June 2013.*

Kim, D., & Kim, H. (2013) "We are still vulnerable to clickjacking attacks: about 99% of Korean websites are dangerous". *Department of Computer Science and Engineering, Sungkyunkwan University, Republic of Korea*

Leitch, S., Warren, M. (2009). "Security Issues Challenging Facebook". *Proceedings of the 7th Australian Information Security Management Conference. 1 (1), pp137-142.*

Lundeen, B., & Alves-Foss, J. (2012). "Practical clickjacking with BeEF In Homeland Security (HST)", *2012 IEEE Conference on Technologies for (pp. 614-619). IEEE.*

Mahmood, S., & Desmedt, Y. (2012). "Your Facebook deactivated friend or a cloaked spy. In Pervasive Computing and Communications Workshops (PERCOM Workshops)", *2012 IEEE International Conference on (pp. 367-373). IEEE.*

Md sazzadur, R, (2012). "Detecting Social Malware and its Ecosystem in Online Social Networks". *Electronic Theses and Dissertations UC Riverside, PhD thesis, pp1-124.*

Mohtasebi, S., & Dehghantanha, A. (2011). "Defusing the Hazards of Social Network Services". *International Journal of Digital Information and Wireless Communications (IJDIWC), 1(2), 504-515.*

Neumann, A., Barnickel, J., & Meyer, U. (2010). "Security and privacy implications of URL shortening services". *In Proceedings of the Workshop on Web 2.0 Security and Privacy.*

Rahman, M. S., Huang, T. K., Madhyastha, H. V., & Faloutsos, M. (2012). Efficient and scalable socware detection in online social networks. In USENIX Security.

Securelist. (2012). "Kaspersky Security Bulletin 2012. The overall statistics for 2012". Available:http://www.securelist.com/en/analysis/204792255/Kaspersky_Security_Bulletin_2012_The_overall_statistics_for_2012#3. Last accessed 20th May, 2013.

Stein, T., Chen, E., & Mangla, K. (2011). "Facebook immune system". *In Proceedings of the 4th Workshop on Social Network Systems (p. 8). ACM.*

Ter Louw, M., & Venkatakrishnan, V. N. (2009). "Blueprint: Robust prevention of cross-site scripting attacks for existing browsers. In Security and Privacy", *2009 30th IEEE Symposium on (pp. 331-346). IEEE.*

Thomas, K., & Nicol, D. M. (2010). "The Koobface botnet and the rise of social malware. In Malicious and Unwanted Software (MALWARE)", *2010 5th International Conference on (pp. 63-70). IEEE.*

Thomas, K., Grier, C., Ma, J., Paxson, V., & Song, D. (2011). Design and evaluation of a real-time URL spam filtering service. In Security and Privacy (SP), 2011 IEEE Symposium on (pp. 447-462). IEEE.

Xu W., Zhang, F., Zhu, S. (2010). "Toward worm detection in online social networks". *In Proceedings of the 25thAnnual Computer Security Applications Conference (ACSAC),*

Yan, G., Chen, G., Eidenbenz, S., & Li, N. (2011). "Malware propagation in online social networks: nature, dynamics, and defence implications". *In Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security (pp. 196-206). ACM.*

Yan, G., Chen, G., Eidenbenz, S., & Li, N. (2011, March). "Malware propagation in online social networks: nature, dynamics, and defense implications". *In Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security (pp. 196-206). ACM.*

Securelist. (2010). "Kaspersky Security Bulletin. Malware Evolution 2010". *Available: http://www.securelist.com/en/analysis/204792161/.* Last accessed 10th Nov 2010.

Countering Threats - a Comprehensive Model for Utilization of Social Media for Security and law Enforcement Authorities

Margarita Jaitner[1] and Harry Kantola[2]
[1] Karlstad University, Karlstad, Sweden
[2] Finnish National Defense University, Helsinki, Finland
m.a.jaitner@gmail.com
harry.kantola@mil.fi

Abstract: The increased adoption of social media has presented security and law enforcement authorities with significant new challenges. For example, the Swedish Security Service (SÄPO) asserts that a large proportion of radicalization takes place in open fora online. Still, approaches to contain social media-driven challenges to security, particularly in democratic societies, remain little explored. Nonetheless, this type of knowledge may become relevant in European countries in the near future: Amongst other factors, the challenging economic situation has resulted in increased public discontent leading to emergence or manifestation of groups that seek to challenge the existing policies by almost any means. Use of social media multiplies the number of vectors that need law enforcement attention. First, a high level of social media adaption allows groups to reach and attract a wider audience. Unlike previously, many groups today consist of a large but very loosely connected network. This lack of cohesion can present a challenge for authorities, to identify emerging key actors and assess threat levels. Second, a high level of mobile web penetration has allowed groups to ad-hoc organize, amend plans and redirect physical activities. Third, the tool social media is as not exclusive to potential perpetrators of unlawful action, but is as well available to law enforcement authorities. Yet, efficient utilization of social media requires a deep understanding of its nature and a well-crafted, comprehensive approach. Acknowledging the broad functionality of social media, as well as its current status in the society, this article describes a model process for security authorities and law enforcement work with social media in general and security services work in particular. The process is cyclic and largely modular. It provides a set of goals and tasks for each stage of a potential event, rather than fixed activities. This allows authorities to adapt the process to individual legal frameworks and organization setups. The approach behind the process is holistic where social media is regarded as both source and destination of information. Ultimately, the process aims at efficiently and effectively mitigating the risk of virtual and physical violence.

Keywords: social media, law enforcement, OSINT

1. Introduction: Social media and security

The rapid development of computer networks during the past decades has left its traces in many, if not all, areas of our lives. In a 2012 report compiled by the Boston Consulting Group analysts have estimated that by 2016 the number of Internet users will globally reach the 3 billion mark (Dean et. al., 2012), a report authored by Scott Bicheno of Strategy Analytics appraised that the number of smartphones in use globally passed the 1 billion mark in the third quarter of 2012 (Reisinger, 2012) The ever-developing connectedness changed, and continues to change, the way we work, the way we entertain ourselves and the way we communicate.

The Internet "has become an ingrained feature of everyday life" (Dean et. al., 2012) in many parts of the world. Sociologist Manuel Castells (2009) calls this the rise of mass self-communication. According to him, mass self-communication possesses certain socio-political power, which he exemplifies with so-called "grassroots"-movements that emerge in the society and grow with the help of shared information within the connectedness and have the potential to influence higher levels in the political system. Indeed, the global interconnectedness "is associated with greater citizen commitment to democratic governance," according to Nisbeth et. al. (2012). But with power to influence and change comes also the power to threaten and destroy, and thus, the potentially harmful aspects of the Internet, and the social media within, began to play a role in political practice, governance and in the related research, as an extension of the physical arena. This relatively new space is also a reflection of large parts of the global society, many of the actors seeking gain or influence in the physical world are today found online - from states to individuals, from organized terror organizations and hooligans and petty criminals.

Recognizing the importance of the new virtual domain, various authorities have shown an increased interest in utilizing cyber in general, and social media in particular, in their work. The virtual reality, however, have tasked organizations to amend their practices to fit its unique features and possibilities. Various methods have been developed and many authorities are nowadays are present in social media and seek to inform (Polisen 2011; Borchers 2011) or take advantage of social media's two-way communication and seek out for public support in

crime investigation (Vancouver Police Department 2011). Arguably, as comprehensive and interconnected the virtual world has become, as comprehensive the approach to work within should be in order to make efficient use of the opportunities as well as to address the pitfalls it offers.

This article seeks to present a process for security and law enforcement authorities to work with social media as a significant part of the virtual arena based on the various opportunities it offers. Because social media can be used for a wide range of purposes, the process was designed as general and flexible as possible in order to be implementable across various authorities with different responsibilities and limitations. Further, the process recognizes the need for the relevant authorities to cooperate and create shared situational awareness particularly during a significant event, in the sense of "shared awareness of a situation," implies that authorities understand a given situation in the same way (Nofi 2000).

Despite its abstractability, process is largely based on various approaches to mitigation of large-scale public action, such as protests or mass violence. Unlike insight into issues closely aligned with national security, the work containing unrest is largely visible and tends to be thoroughly scrutinized in population, media, as well as academia. The recent years' events, such as the riots in the UK in 2010 (Gross 2011), the events of the Arab Spring (Zuckerman 2011), as well as the Russian election-related protests in 2011/2012 (Jaitner 2013) have provided a rich empirical base of activities within social media. Arguably, the mechanisms derived from this empirical base are of use beyond addressing riots or mass violence. Consequently, the process for work in social media can be used for various scopes of attention: those tasked with mitigation of protests and mass action may need to pay attention to the sentiment of the general population, while another authority may focus their work on a particular group. In the scope of the process, social media signifies any space in the cyber world that provides for exchange of information. Depending on area of implementation, it may, however, be useful to apply a narrow definition.

2. Social media as a source of operational data

Arguably, the key to social media is the information within, and for those who wish to utilize social media the operationability of information is of interest. Generally, information can be grouped by the following criteria:

- Content in form of text, sounds, or pictures, and its contextual relevance
- Contextual information, for example in form of user profiles and their logical networks
- Intensity, which encompasses frequency of use or volume of data transmitted

(Beatty, 2006)

Different platforms in the physical world as well as in cyberspace provide for different possibilities to extract such information. In order to be able to work with any platform there needs to be knowledge of what information can be extracted and how (Beynon-Davies, 2002:47-49). Likewise, before attempting to create a process for utilization of social media we need to know how information is presented within different platforms. As earlier mentioned, there is a vast majority of websites and applications that can be considered social media. Many of these have distinct functionality, some encompass elaborate user profiles, such as Facebook or LinkedIn, while other lack such profiles and rather focus on content. Examples for the latter are various news outlets that provide for commentary function (van Dijck, 2013:8). Despite this inhomogeneity of provided information, similar sets of data can be extracted.

To exemplify data extraction from social media a Twitter snapshot (Figure 1) can be used. Twitter in this case is representative for many other types of social media, with respect to differences in functionality, or alternatively can demonstrate a cross-platform inquiry.

Based on the different categories of data mentioned above, the following questions are asked:

1. What is the content?

The content-based inquiry indicates that twitterers in Eastern Europe are, amongst other topics, talking about "euromaidan", "Grushevskogo" (Грушевского) and "Berkut" (Беркут).

Figure 1: Twitter trend map, Ukraine, 22. Jan. 2014

2. What is the context?

The trend map indicates that people are connecting with the user "@s_vakarchuk" in Twitter-based conversations. Further, the geographical location of various tweets can be identified.

3. What is the intensity?

In this particular map trending topics on Twitter are pictured in different sizes depending on the volume of activity. In this particular case a high level of activity near Lvov and Kiev is observable. Ideally such indicators are mapped to numbers of tweets.

Combining the results of the inquiries we can note that a lot of people in Eastern Europe are tweeting about "Euromaidan" and are engaging in conversations with the user @s_vakarchuk. Based on data that results from such inquiries, metrics can be established for the recognized situational awareness, the "normal situation". These metrics provide for answers to following questions: How much social media activity is there normally? Which are the established groups in the society? What are the usual topics? How often does focus from one topic to another change? Where are the usual locations? This includes a reflection on usual volume of activity in different situations, such as before festivities or particular events: during elections or large cultural events activity volume may rise significantly. Also an estimation of "normal topics" in relation to current news and trends is necessary.

Several solutions for creating situational awareness within social media are currently available on the market, ranging from built-in analysis capability in social media over Google analytics and the like to advanced solutions for professional intelligence and law enforcement use, like Raytheon RIOT or GEOCOP. Depending on the individual solution, analytics may focus on overall activity, or activities within certain groups. In short, social media can provide three general types of data that in different combinations can be used to indicate public or group activity. Additional types can be extracted under particular circumstances, such as location in association with gatherings. Other tasks may require extraction of other information as a separate data type. Moreover, it is technically possible to compile and analyze the available data in a meaningful way that can be used to support authorities in their work. Taking these possibilities as a starting point, the opportunities to utilize this data will be examined in the following chapter.

3. Reconnaissance

Observations during the Russian election-related protests of 2011/2012 provide a highly structured use of various platforms: While Twitter was intensely used during the actual protests, a lot of planning was conducted in Facebook and VKontakte "Group" and "Event" pages and blogs were widely used to nurture the political discourse (Jaitner, 2013). In UK riots, on the other hand, BlackBerry Messaging was evidently the platform of choice for communication between the rioters. At the same time, vivid riot-related discourse on amongst others Twitter and Facebook lead authorities to believe these platforms were the culprits (Ball &

Brown, 2011). This demonstrates the importance of an accurate overview over the situation including awareness of different platforms, their possible purposes, and their users. While a significant amount of such awareness originates in monitoring itself, it is important to map and assess the terrain beforehand in a reconnaissance activity.

Reconnaissance is the bedrock of any further efforts within social media. At a very basic level such activity results in awareness for a further "dimension" in which public discourse is conducted, as well as a basic scheme of platforms and their functions, which then has to be analyzed as to how use of these platforms can result in hazardous situations. Also, metrics for the "normal" have to be established in terms of activity levels tied to various temporal and event-related factors. This constitutes creation of a shared situational awareness, which provides a base for development of further activities. A wide range of procedures for reconnaissance is possible, from manual collection and analysis of data to computer assisted mapping. Given the complexity of social media an interdisciplinary approach to reconnaissance is necessary. For example, a rise to a certain level of intensity in one cultural area can be caused by a festivity, while being a sign of an emerging event in another. Therefore, those who are appointed to work with the data have to be familiar with what is to be regarded as normal at different times.

4. Monitoring

In many instances authorities with a public security mandate already utilize the information available in social media in monitoring activities. Examples can be found in

Norwegian Police Security Service's (Norwegian: Politiets sikkerhetstjeneste - PST) report for 2011, where PST points out the organizations concern with "the use of social media as a contact and organizing platform" for potentially dangerous groups. The Swedish Security Service – Säkerhetspolisen – lists "the Internet" as an open source for information retrieval in the authorities' report for year 2012, stating that "recruitment, radicalization and financing is often conducted via social contacts, as well as lectures and propaganda activities, amongst other on the Internet" (Säkerhetspolisen, 2013:22) in regard to radical Islamist terrorism. Findings of an ICSR study on far-right extremism in Europe (Meleagrou-Hitchens & Brun, 2013) allow assuming that the situation is similar regarding many other potentially radicalized groups.

Solely observational and non-interceptive activity within social media closely resembles monitoring activities in other security or even business-related areas. It is conducted in order to raise awareness for the currently important issues; it is also crucial for utilization of social media in any further activity - successfully dealing with an event requires knowledge regarding its various aspects.

At a very basic level, this activity is conducted as a part of day-to-day business at various levels in the hierarchy of any organization: Personnel in different roles and in different departments, consciously or unconsciously "monitors" their environment and utilizes this knowledge in one way or another in their work. Unstructured observations are conducted for example via personnel's own personal social media experiences, media reporting on social media, collective shared knowledge, and rumors. This type of information helps individuals in their day-to-day work as well as in exceptional situations and during events. The problem, however is, that such observations might be wrong, they might deliver insufficient background information, lack of agreed interpretation and analysis, also important features may not draw enough attention to be noticed.

In this activity all three types of information are collected. Relationships between groups have to be established and evaluated regarding their potential threat. Moreover, a normal level of intensity has to be established. Outside event various types of content will be present, amongst them types that may be alarming. Therefore it is important to establish, what level of activity within this particular content constitutes a normal. Similarly, it may be useful to map groups to particular content and intensity of activity. In this way a more precise situational awareness is created that is sensitive for smaller changes, such as sudden shifts in intensity amongst a particular group that may indicate distress or imminent changes. These changes can be treated as potential triggers, which will be needed for identification of a possible upcoming event. It is important to take regard to the situation in the physical world when identifying such triggers.

During monitoring content, relationships, location and intensity information is collected. A constant comparison with "the normal" within the various types and subtypes of information is necessary at this point.

This can also include particular styles and patterns of communication amongst subgroups that are of interest. Depending on the overall situation, authorities may chose to focus on one primary type and treat other information as secondary. For instance, if groups that potentially can threaten public order are identified, it can make sense to monitor groups in regard to content and intensity. Content and particularly intensity in their turn can be linked to a particular geographic location. For example, higher-than-normal frequency in postings made by a hostile group at a particular gathering place can raise concerns about a possible unsanctioned mass gathering (Figure 2).

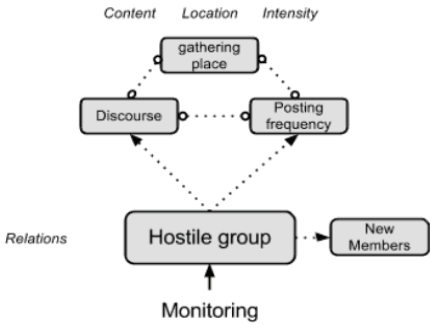

Figure 2: Relationship-based monitoring

In certain societies it may be important to focus on the content, rather than on groups. In this case content becomes the primary subject of monitoring and is put in context by intensity and relationships. On the other hand if it is known that overall activity remains fairly constant over time, authorities may choose to treat intensity as the primary type of information.

5. Anomaly detection

With regard to the established recognized situational awareness, any anomaly would be identified as an event that needs further attention. This may be an unusual amount of activity in a certain geographic location, unexpected discussion between groups or sudden change in rhetoric and discourse. Even sudden formation or mobilization of groups deserves attention. Because anomalies can present in very different ways across the various platforms (Jaitner, 2013), it is essential to recognize the differences between the various types of social media. A successful case of event recognition appears to be the preventive measures taken by schools and police following the first day of student riots in Gothenburg in 2012. There, schools and authorities seem to have recognized an imminent mobilization of a large group in social media and responded with school closures and police presence in affected areas. (Svahn & Bolling, 2012)

Identification of unexpected activity within social media is generally a sub activity to monitoring, albeit an important one. Anomalies can present over time or arise suddenly consecutive to an emotionally charged planned or unplanned event. In this activity all three information types can be of importance. Particular triggers identified during reconnaissance should be in focus. Generally triggers can present in form of rise in overall activity (intensity), sudden aggravated communications between disagreeing groups (relationships), or a particular discourse (content). Pohl et. al. (2012) argue that sub-events can be automatically detected during an overall emergency situation using multimedia meta data clustered approach. Reasonably, a parallel can be drawn to normalcy, where events would be identified. Implementation of such automatisms is certainly desirable. The sooner an event is identified, the more time authorities will have to plan and initiate following activities and the better prepared the organization will be once the event becomes a hazardous. However, even though imminence of an event naturally draws increased attention to a particular sector or group, overall monitoring should continue.

6. Information activities

Showing presence for deterrence of disorderly behavior and crime prevention as well as supplying citizens with information on how to avoid or handle potentially dangerous situations or to report a crime is widely recognized as an essential part of police work (Polisen 2011) Because today's society is not limited to the physical world, authorities recognize the need to extend their presence into the cyberspace (Polisen 2011, Vancouver Police Department, 2011:56).

Generally two approaches can be observed: one-way communication where authorities seek to inform the population and two-way communication that encompasses maintaining a conversation, including receiving feedback or other information that may assist authorities in their work (Vancouver Police Department, 2011; YLE 2012). In this context, information can be sent to the general population or to individual groups with a particular intensity. Alternatively, authorities can engage in conversations with the general population or particular groups. Any significant information that is retrieved during informational activities should be treated as subject of monitoring, because triggers can present during informational activities. Identification of information that fuels the rage, as well as identification of trends, provides an important base for further operations. Research suggests that hazard-related (Pohl et. al. 2012) and emotionally charged (Berger & Milkman 2012; Stieglitz & Dang-Xuan 2012; Spiro et. al. 2012) information travels faster through social media than other types of information. From that point, various approaches can be taken to reduce pressure.

7. Reducing pressure

Reducing pressure is a sub activity to information activities and seeks to mitigate or even avert an imminent event. Methods for reducing pressure encompass, but are not limited to: active participation in discussions within a distressed group, disproving false rumors, diversification of online discourse, disinformation and limitation of access to media used to "fuel" the distress (Jaitner 2013). Any method has to be conducted with consideration for behavioral, cultural, and socio-economical factors, as well as the judicial situation.

Although it may be difficult to pinpoint the exact effect of such activity, research shows that public (mass media) interest, or lack thereof, affects the course of group-driven events (Myers 2000). Behavior that can be perceived as consciously misleading or ineffective might result in loss of legitimacy and support for authorities amongst groups that previously showed disapproval of protesting or rioting. This activity can be embedded within overall informational activities, unless conducted in a more coercive manner.

8. Supporting the physical world

Naturally, any information that can enable planners of personnel on site should be taken advantage of. Particularly geolocation-enabled information can be of importance, but also other extractable information alone or in combination can deliver valuable decision support for personnel tasked with dealing with the physical event: New groups can form ad hoc during an event, formerly cohesive groups can disintegrate and become hostile within. Further, geographical context paired with intensity can help locate hotspots and allow estimating number of personnel needed on the ground, while content can give leads on how the event might unfold.

9. Supporting the judicial system

Information in social media can also be gathered for judicial support. Here the purpose ranges from collecting evidence to identifying persons of interest. Authorities can make use of the wide reach and fast dissemination within social media in order to attract public assistance. Requests for assistance are relatively widespread across various social media and even authorities have made use of this method. In the wake of a riot in Vancouver, Canada, the local police department requested citizens to submit evidence and to help to identify perpetrators (Vancouver PD 2011).

The phenomenon of self-motivated "citizen detectives" could also be observed shortly after the Boston Marathon Bombings in 2013. Social media crowd posted and analyzed videos and photos on the events and discussed possible suspects in social media resulting in what was later dubbed the "Witch hunt" - the crowd had identified evidently innocent persons as possible perpetrators, which had been quickly picked up by large traditional media and reported globally. (Ingram 2013). This case provides three insights: First, the potential strength of popular self-motivation for what is perceived to be the right cause. Second, it demonstrated the ability of social media users to gather (and even analyze) massive amounts of data in a short time. The third insight is that the online discourse was not limited to attempts to find the perpetrators but contained information how people could assist victims of the bombings - "from links to Google's Person Finder and the Red Cross help line to information on where to pick up bags left at the scene, or airlines who had changed their policies on cancelling flights as a result of the attacks," writes Matthew Ingram (2013). Arguably, these aspects of user activity can be utilized by authorities during and post-event if deemed necessary. At the same time, the strong self-directed crowd participation implies a risk on its own and thus may need to be mitigated.

10. The process

In order to create an implementable framework the individual activities are put into a process model. Before personnel can work within social media, shared situational awareness has to be developed. The activity of reconnaissance then transforms into monitoring in order to be able to identify anomalies. When atypical activity is identified and the analysis thereof indicates possible distress within a group, efforts to reduce pressure can be carried out and continued during the actual event. At onset of the event, data collected in social media can then support decisions on whether or not to deploy ground personnel, as well as estimate personnel requirements at different locations. At the same time evidence collection can begin. Local laws may dictate that authorities can first start collecting evidence whenever criminal behavior is observed. Evidence collection can continue during normalization after the event. After the event it is likewise necessary to investigate whether any significant changes have occurred, making a repeated reconnaissance activity necessary. The summarized model is presented in Figure 5.

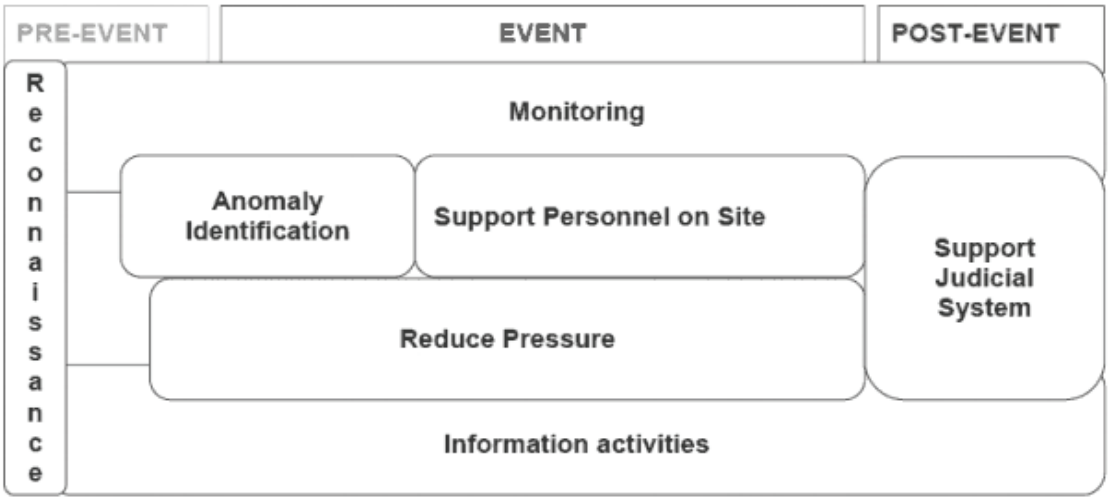

Figure 3: Process for security and law enforcement work with social media

11. Conclusions - what about applicability?

When implementing the process, authorities in question may want to incorporate already established methods for work within social media, which is likely to lead to these methods being further developed to suit more than one need. The process can be applied with varying levels of coerciveness and intrusion. Depending on the overall setting with regard to culture as well as national judicial constraints different levels of participation, from intrusion by infiltration to observation of openly available sources, can be interpreted in different ways in terms of coercion. In a society that is used to a high level of oversight by state authorities intrusion into closed fora may not lead to loss of soft power, on the contrary, this type of activity amongst potentially dangerous groups may be expected by the broad society. In cultures that are built upon transparency, on the other hand, intrusive activity may be regarded as unacceptable. Generally, a low level of coerciveness by authorities is assumed for application of the model. Instead it is designed to enhance their soft power through active participation in the discourse and positive presence in the social media.

Availability of adequately trained personnel, funding restrictions or overall working strategy may hinder or disallow certain activities within the process from being executed. However, as long as an accurate situational awareness is provided, authorities may choose to conduct individual or a set of activities within the model. For example, in some cases events will develop with a speed that does not allow for timely event recognition, thus making proactivity impossible. Shared situational awareness in its turn can be developed through sole monitoring over a certain period of time. A reality check, and a direct connection to the physical world are provided particularly while supporting personnel that acts in the physical world. At that point, ‘virtual authorities’ have to interact with ‘physical authorities’. However, the ‘virtual’ may misunderstand the ‘physical’, and the ‘physical’ may not regard the ‘virtual’ with the needed seriousness. This is not per se a disadvantage of the process itself but rather an organizational, or structural, problem to be aware of. It also enhances the necessity to tightly integrate the work in the cyberspace with work in the physical world - which in itself can provide a chance to bring about a greater understanding of the virtual, but no less real, world.

References

Ball, J., Brown, S. (2011). Why BlackBerry Messenger was rioters' communication method of choice. The Guardian, December 20. Retrieved May 20, 2013 from http://www.guardian.co.uk/uk/2011/dec/07/bbm-rioters-communication-method-choice

Beatty, A. (2006). Studying Media Effects on Children and Youth: Improving Methods and Measures, Workshop Summary. National Academies Press.

Berger, J., & Milkman, K. (2012). What makes online content viral? Journal of Marketing Research, 49, 192–205.

Borchers, D. (2011) Polizei in Sozialen Netzwerken: Verdeckte Ermitter online. Frankfurter Allgemeine Zeitung, August 04. Retrieved May 20, 2013 from http://www.faz.net/aktuell/technik-motor/computer-internet/polizei-in-sozialen-netzwerken-verdeckte-ermittler-online-11114160.html

Bynon-Davies, P. (2002). Information Systems: An Introduction to Informatics in Organisations. Hampshire: Palgrave.

Castells, M. (2009). Communication power. OUP Oxford.

Dean, D., DiGrande, S., Field, D. & Zwillenberg, P. (2012). The Connected World:The Digital Manifesto: How Companies and Countries Can Win in the Digital Economy. BCG perspectives, JANUARY 27, 2012. Retrieved May 20, 2013 from https://www.bcgperspectives.com/content/articles/growth_innovation_connected_world_digital_manifesto/

van Dijck, J (2013) The Culture of Connectivity: A Critical History of Social Media. Oxford: Oxford University Press.

Gross, D. (2011). British PM proposes social media ban for rioters. CNN, August 12, 2011. Retrieved May 20, 2013 from http://edition.cnn.com/2011/TECH/social.media/08/11/london.riots.social.media/index.html

Ingram, M. (2013). Three things that Reddit did right during the Boston bombings and why that matters, April 23. Retrieved May 20, 2013 from http://paidcontent.org/2013/04/23/three-things-that-reddit-did-right-during-the-boston-bombings-and-why-that-matters/

Jaitner, M. (2013). Exercising Power in Social Media. In Rantapelkonen, J., & Salminen, M. (eds.) (2013). The fog of cyber defence. Julkaisusarja 2. Artikkelikokoelma n: o 10. pp. 57ff

Meleagrou- Hitchens, A., Brun, H. (2013) A Neo-Nationalist Network: The English Defence League and Europe's Counter-Jihad Movement. The International Center for Study of Radicalization and Political Violence. Retrieved May 20, 2013 from http://icsr.info/wp-content/uploads/2013/03/ICSR-ECJM-Report_Online.pdf

Myers, D. J. (2000). The Diffusion of Collective Violence: Infectiousness, Susceptibility, and Mass Media Networks. American Journal of Sociology, 106(1), 173-208.

Nisbet, E. C., Stoycheff, E. and Pearce, K. E. (2012), Internet Use and Democratic Demands: A Multinational, Multilevel Model of Internet Use and Citizen Attitudes About Democracy. Journal of Communication, 62: 249–265

Nofi, Albert A. (2000) Defining and Measuring Shared Situational Awareness, Center for Naval Analyses, Virginia, USA. Retrieved May 20, 2013 from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA390136

Pohl, D., Bouchachia, A., & Hellwagner, H. (2012, April). Automatic sub-event detection in emergency management using social media. In Proceedings of the 21st international conference companion on World Wide Web (pp. 683-686). ACM.

Polisen, (2011). På plats i sociala medier. Tidningen Sambandet, March 15. Retrieved May 20, 2013 from http://www.polisen.se/Om-polisen/lan/St/op/Polisen-i-Stockholms-lan/Sambandet/Sambandet/Oktober-2008/Pa-plats-i-sociala-medier/

Politiets sikkerhetstjeneste (n.d.). PST1011: Kontraterror. Report for 2011. Retrieved May 20, 2013 from http://www.pst1011.no/2-kontraterror/

Reisinger, D. (2012). Worldwide smartphone use base hits 1 billion. CNET, October 17. Retrieved May 20, 2013 from http://news.cnet.com/8301-1035_3-57534132-94/worldwide-smartphone-user-base-hits-1-billion/

Säkerhetspolisen (2013). Säkerhetspolisen 2012. Årsbok, April. Retrieved May 20, 2013 from http://www.sakerhetspolisen.se/download/18.3fbb83313df609c410800025/Sapoarsbok2012extwebb.pdf

Stieglitz, S. & Dang-Xuan, L. (2012). Impact and diffusion of sentiment in political communication: An empirical analysis of public political Facebook pages. Proceedings of the 20th European Conference on Information Systems (ECIS), Barcelona, Spain.

Spiro, E. S., Irvine, C. A., DuBois, C. L., & Butts, C. T. (2012). Waiting for a Retweet: Modeling Waiting Times in Information Propagation. Stanford Network Analysis Project. Retrieved May 20, 2013 from http://snap.stanford.edu/social2012/papers/spiro-dubois-butts.pdf .

Svahn, C., Bolling, A. (2012). Spänt vid Göteborgsgymnasium. Dagens Nyheter: Skoluppploppet i Göteborg 2012, December 19. Retrieved May 03, 2013 from http://www.dn.se/nyheter/sverige/framtidsgymnasiet-stanger-efter-hot

Vancouver Police Department, (2011) Vancouver PoliceDepartment 2011 Stanley Cup Riot Review. Retrieved May 20, 2013 from http://vancouver.ca/police/assets/pdf/reports-policies/vpd-riot-review.pdf

YLE (2013) Spotlight: Infiltratören. February 20. Retrieved May 20, 2013 from http://svenska.yle.fi/artikel/2013/02/20/spotlight-infiltratoren

Zuckerman, E. (2011). The First Twitter Revolution? Foreign Policy, January 14, 2011. Retrieved May 20, 2013 from http://www.foreignpolicy.com/articles/2011/01/14/the_first_twitter_revolution

Potential Cyber Warfare Capabilities of Major Technology Vendors

Audun Jøsang
University of Oslo, Norway
josang@ifi.uio.no

Abstract: The Snowden revelations of 2013 gave the public an insight into the cyber surveillance capabilities of the NSA. These capabilities depend largely on the collaboration from some of the major providers of online services and from social network providers such as Facebook and Google. A topic that has not received much focus are the potential cyber operations capabilities and advantages that governments could obtain through the collaboration with major domestic technology companies, such as the software vendor Microsoft and hardware vendor Intel in the USA, or the telecommunication network equipment vendor Huawei in China. While the Snowden revelations focused on surveillance capabilities, the potential collaboration with software and hardware vendors could give tremendous offensive cyber operations capabilities. This paper discusses the cyber warfare capabilities that could be gained through the collaboration with major software and hardware vendors as well as with IT service providers, and explores the political conditions for this type of collaboration to take place and the possible consequences of such scenarios.

Keywords: cyber operations, industry collaboration, Snowden revelations, operating system, hardware

1. Introduction

The term 'cyber operations' is often used to denote the set of methods used for cyber warfare. Surveillance, defense and attack in cyberspace are general elements of cyber operations that in turn can be decomposed into more specific elements.

In its policy document (White House, 2012) the US administration defines cyber operations as consisting of: *"Cyber Collection, Defensive Cyber Effects Operations (DCEO) and Offensive Cyber Effects Operations (OCEO)"*. To our knowledge the US administration is the only one in the world that officially has a cyber operations policy clearly articulated in this way. Obviously many other national administrations also practice cyber operations, but policy makers in nation states can reason that there is no need to publish their cyber operations policies because these activities are relatively easy to hide from the public. We find laudable the US administration's publication of their cyber operations policy because it reduces confusion and contributes to transparency in the cyber warfare landscape, which in turn is beneficial for political stability.

According to the same policy document, surveillance through computer networks is called 'Cyber Collection' which is defined as: *"Operations and related programs or activities conducted […] for the primary purpose of collecting intelligence - including information that can be used for future operations - from computers, information or communications systems, or networks with the intent to remain undetected. Cyber collection entails accessing a computer, information system, or network without authorization from the owner or operator of that computer, information system, or network or from a party to a communication or by exceeding authorized access. […]."*

Methods for surveillance of global communication networks are many and diverse, ranging from interception of communication passing through cables or through the air, to obtaining data stored in databases and being processed within computer hosts. Data can often be intercepted without the authorization and collaboration of the network or host owners as described in the cyber collection policy cited above, but surveillance capabilities can be significantly amplified with their collaboration. The revelations from late 2012 and throughout 2013 by Edward Snowden about NSA's surveillance capabilities clearly documented that the NSA (US National Security Agency) obtained active collaboration from several prominent companies in order to enhance NSA's surveillance capabilities. This was done e.g. by having arrangements with companies like Apple, Google and Facebook for obtaining access to customers' data and profiles (Greenwald and MacAskill, 2013), which in fact is legal according to the USA PATRIOT Act (107th US Congress,2001).

It has also been revealed that NSA paid US$ 10 million to RSA, a major security technology vendor, to include a backdoor in one of their cryptographic products, thereby enabling the NSA to read traffic encrypted with that product (Menn, 2013). This form of surveillance based on active and consenting collaboration with private companies can be considered to be a form of authorized access and therefore seems to extend the scope of the above cited cyber collection policy. However, when assuming that the targets of surveillance are the

customers of the collaborating parties, not the collaborating companies, then the surveillance falls under the definition of unauthorized 'cyber collection' as cited above. In fact, the collaborating companies become part of the surveillance operations targeted at their own customers. In this scenario it is thus meaningful to distinguish between 3 separate parties: 1) the cyber operations agent, 2) the collaborator, and 3) the target of cyber operations.

Defensive and offensive cyber operations can in general be conducted with or without the collaboration of external parties, and similarly to cyber collection, the potential for defensive and offensive cyber operations increases significantly with the assistance of collaborating parties having specific skills or who control specific technologies and ICT infrastructures. In the sections below we explore various methods that potentially could be used for cyber operations and that have not been mentioned in the literature. These methods go far beyond those described through the Snowden revelations.

Collaborating parties in cyber operations can be public or private companies and organizations, or they can be other nation states. In general, the collaborator can be seen as an intermediary between the cyber operations agent and the target. It is normally assumed that there is a friendly relationship between the cyber operations agent and the collaborating party; otherwise the collaborating party should more precisely be seen as a target of cyber operations. It is thus the nature of the business relationship or the political alliance between the cyber operations agent and an involved party that determines whether the involved party is a collaborator or a target of cyber operations.

Various forms of collaboration are possible, ranging from full voluntary and active participation in the cyber operations, to forced and involuntary participation. Although there is a continuum of types of collaboration we find it meaningful to distinguish between active, passive and forced collaboration that we define as follows.

- *Active collaboration* is when representatives of the collaborating party take voluntary conscious actions in order to assist in cyber operations. This is for example the case when Facebook transmits user profiles to NSA, or gives NSA technical access to the data at their discretion.
- *Passive collaboration* is when representatives of the collaborating party voluntarily and consciously do not exercise their right to take actions that could have hindered cyber operations. This is for example the case when RSA discovers a vulnerability in one of their security products which they know can be exploited by NSA, and they do not attempt remove that vulnerability.
- *Forced collaboration* is when national authorities use pressure or a legal right to force the collaborating party to contribute to cyber operations. Forced collaboration is possible when applicable laws dictate the right to obtain information from collaborating parties. This was for example the case when the U.S. government in 2013 obtained a search warrant demanding that Edward Snowden's e-mail provider, Lavabit, turn over the private SSL keys that protected all web traffic to the site. Lavabit offered paying customers a secure email service that offered the highest possible level of privacy. When faced with the search warrant, Lavabit decided to stop operations rather than betray its own customers (Poulse, 2013).

The form in which collaboration is practiced in cyber operations can have implications for legal liability, ethical standing and public relations of the collaborating party. A company that grants a cyber-operations agent access to its resources could be the target of legal action, public criticism and could suffer damage to its reputation should this collaboration be publicly known. As an attempt to restore their lost reputation, cyber operations collaborators (e.g. Google, Apple, Facebook) identified by the Snowden revelations have argued for legal changes regarding this form of cyber operations (Roberts and Kiss, 2013). IBM issued a statement saying that it will take steps to challenge any order by the US government to hand over data for surveillance purposes (Weber, 2014), but so far there are no cases where this has happened.

The revelation of massive collaboration between NSA and online social network providers has probably resulted in withdrawal by target groups from these networks precisely because they want to avoid being spied upon. From the perspective of NSA's cyber operations it is therefore likely that the Snowden revelations have reduced the value of collaboration with these social network providers. Said bluntly, if terrorists used Facebook before the Snowden revelations, most likely they do not use Facebook anymore. The secrecy of specific cyber operations is thus crucial for their effectiveness. Another negative side-effect is of course the lost business suffered by the collaborating companies. They would probably have a credible legal argument for

financial compensation of damage to their reputation and resulting losses due to their collaboration with NSA, because they can claim that it unfairly damages their business.

Cyber operations (DCEO, ODEO and Cyber Collection) can in general be exercised with or without the collaboration of external parties. Cyber operations exercised without any collaborating parties will resemble typical cyber attacks that require the exploitation of vulnerabilities in systems and networks, or the manipulation of individuals through social engineering. Exploited and discovered vulnerabilities are constantly being removed or reduced by system vendors and owners. In order to maintain cyber operations capabilities, there is thus a need for continuously developing new attack methods and discovering new vulnerabilities. New and unknown vulnerabilities are constantly being introduced through technology innovation and product development, where most vulnerabilities are introduced by accident through poor security design, but some can be introduced on purpose e.g. as part of cyber operations.

Although old vulnerabilities get fixed, the rate of eradication is eclipsed by the rate of introduction of new vulnerabilities. Numerous players invest considerable resources in trying to discover these vulnerabilities, either to be used in own cyber operations or to be sold and used for profit. For the latter purpose there are unregulated markets for trading security vulnerabilities and zero-day exploits, (Ablon, Libicki, and Golay, 2014). In theory a national cyber operations player could obtain the needed exploits from such markets. However, these markets represent relatively unreliable sources of exploits because the buyer can never be 100% certain that the seller does not offer the same vulnerability to the buyer's opponents. National states with an offensive cyber operations program therefore needs to also have the necessary capabilities and resources for 'vulnerability mining' based on using massive computational resources and analytical methods such as fuzzing which is a form of brute force vulnerability discovery (Sutton, Greene and Amini, 2007). Software vendors, criminals and governments alike engage in fuzzing. Software vendors do it to remove software bugs in their products; criminals do it to develop exploits that can generate profit, and governments do it to sustain their cyber operations capabilities.

2. Potential cyber operations capabilities through collaboration with technology vendors

NSA's collaboration arrangement with Google, Facebook and other social network providers is mainly used for surveillance purposes. It is interesting to think further and reflect on the cyber operations capabilities that can be obtained with the collaboration from technology vendors such as Microsoft, Apple, Intel and AMD, Dell and HP. These US companies sell products with dominant global market shares, so that they could be harnessed as potent cyberwarfare tools for the US government. Similarly, telecommunication network equipment produced by Huawei and ZTE could potentially be used by the Chinese government in their cyber operations. This possibility has led some governments to ban their products in specific market segments considered part of their national CII (Critical Information Infrastructure), e.g. in Australia (McDuling, 2012) and in the USA (Rogers and Ruppersberger, 2012). The fact that some western governments have excluded Chinese manufacturer Huawei from critical market segments shows that governments actually consider cyber operations collaboration with technology vendors as a realistic scenario.

Reduced competition with resulting increased costs is an unavodable negative side-effect of excluding major vendors from offering components when building critical ICT infrastructures. In some Western countries such as Norway where there also has been a debate about the risks of using Huawei components in critical national telecommunication networks, the argument about increased costs was one reason for not excluding Huawei.

The sections below explore various cyber operations methods that potentially can be harnessed with the collaboration of technology providers in the respective domains of 1) OS (Operating Systems), 2) CPU (Central Processing Unit) hardware, 2) computer systems, and 4) Internet clouds.

2.1 Collaboration with operating system vendors

Microsoft and Apple develop operating system software for the PC and Mac respectively. Linux OS distributions are developed by multiple vendors, where the code is often open-source. Automated online software update and patching is an integral part and an important feature of these operating systems, which contributes to their security and robustness. Operating systems are large complex software products that therefore necessarily contain high numbers of bugs and vulnerabilities. As soon as a vulnerability has been discovered it can be exploited by people knowing about it. Exploits for known vulnerabilities are being

implemented in automated tools which makes exploitation of those vulnerabilities relatively easy even for non-experts. Without the ability to continuously fix such problems operating systems would be hopelessly unreliable and vulnerable to attacks.

The mechanism for updating and patching is based on programming the client operating system to connect to a specific server owned by the vendor on a regular basis, e.g. daily or weekly. The server checks the current status of the software, and triggers the downloading of missing updates and patches to the client operating system. The updating process is totally controlled by the vendor of the operating system and is mostly transparent to the user. The same updating process can be used to make arbitrary changes to the operating system, so the operating system is in practice totally controlled by the vendor. The vendor of a particular operating system can thus install any application on any computer running its operating system. Given that the Windows OS is installed on approximately 90% of all desktop and laptop computers worldwide the possibility of total control of the majority of client computers worldwide by Microsoft is real. The potential for cyber operations with Microsoft as collaborating party is thus massive. The possibilities range from cyber collection as in surveillance to offensive cyber effects operations as when actively controlling individual computers like a bot machine.

It is unknown to the public whether operating system vendors like Microsoft collaborate with NSA for surveillance and control of individual computers. While the extent of surveillance through Google, Facebook and Apple was shocking to many, collaboration with e.g. Microsoft for controlling individual computers or clusters of computers would have been even more disturbing to the public. There is thus a significant risk for an OS vendor to collaborate in cyber operations. In the case of Microsoft with almost monopolistic market dominance, the consequence of revealing cyber operations collaboration to the public would be hard to predict. On the one hand, cyber operations victims would be reluctant to continue using Windows, but on the other hand there would in many cases be no real alternative. However severe the negative market consequences of cyber collaboration with OS vendors, if a government perceives the threat of e.g. a catastrophic terrorist attack to be sufficiently high, and the only way to uncover or stop the plot is e.g. through controlling computers through OS update, it would arguably be rational to do precisely that.

2.2 Collaboration with CPU and microchip vendors

The layer underneath the OS is typically the CPU hardware. The two main CPU vendors for servers and personal computers are Intel and AMD who share the global market between them and who compete on innovation. Close cooperation between the CPU and operating system vendors leads to rapid innovation and steady increase in performance. This cooperation also results in significant security innovation e.g. to prevent buffer overflow attacks, and to enable secure virtualization of multiple operating systems running simultaneously on the same computer.

It would be possible to implement special triggers that could change the behavior of CPUs as an element of cyber operations to take control of computers. Knowledge of such triggers would enable the control of a computer independently of the operating system running on the computer.

Similar business risks to those mentioned in Sec.2.1 above would exist in case Intel and AMD engage in this form of cyber operations collaboration with NSA. These companies could suffer severe loss of trust from the market should this form of collaboration be revealed, but since they have a monopolistic position it would be difficult for targets to switch to alternative vendors.

Trusted computing is the term used when security functions are based on specialized security hardware. The TPM (Trusted Computing Platform) is a specialized security chip that is integrated on the motherboard of most modern computers shipped today. The role of the TPM is mainly to verify the integrity of low level software such as BIOS and kernel. This can e.g. be used to detect malware infection on the computer. The TPM also has the potential of preventing owners from making certain configuration changes to their own computers, because the TPM dictates certain parts of the configuration and can also dictate which BIOS and OS kernel that can be loaded into memory. In case an exploit has been built into the BIOS or kernel of a computer equipped with TPM, then it might be impossible for the owner to remove those exploits.

2.3 Collaboration with system vendors

During 2013 it was revealed that the NSA disposes of an extensive set of exploits against a large number of IT products (Appelbaum, Horchert and Stöcker, 2013). The exploits are e.g. based on injecting code into the BIOS or kernel of operating systems of computers which can activate network communication, microphone or camera, or it can be tiny hardware chips that can be integrated into systems such as mobile phones, personal computers, servers, routers and firewalls, from major technology vendors such as Dell, HP, Cisco and Apple. It is important to notice that these exploits do not necessarily imply direct collaboration with these technology vendors. It is theoretically possible that the technology vendors have not played any role in developing these exploits, but it is also possible that they assisted in their development, or actually developed the exploits themselves. The revelation of exploits does not say anything about their origin. So even if a vendor participated in the development of an exploits against its own products, and the exploit, but not its origin, is revealed to the public, then it does not necessarily result in lost trust from the market.

There is a huge logistic challenge when executing an exploit like integrating a specialized hardware chip on the motherboard of a computer for cyber operations purposes. It would require physical access to the computer, preferably before it is shipped to the target party. It has been suggested that it can be done during shipping with the collaboration of postal or courier service operators (Appelbaum, and Horchert, and Stöcker, 2013). It would probably be much simpler when done at the assembly line of companies like Dell or HP, or during customization by the local distributors, but the logistic challenge would be significant even here.

2.4 Collaboration with cloud providers

Another domain of government-industry collaboration with great potential for cyber operations could be in the cloud services industry, whereby a cloud service provider grants its national government passive or active access to customer data and processing. Severe negative consequences for the industry player would of course be the result should this form of collaboration become publicly known. Primarily it would have resulted in significant loss of trust and lost business for the collaborating party because targets for cyber operations would switch to alternative cloud providers.

Online social networks can be considered as a form of cloud service, so for surveillance purposes cyber collection collaboration with cloud providers would be very similar. The topic of offensive cyber operations with a cloud provider is difficult to describe or categorize, and because there is a broad range of possibilities the potential is unclear. The potential for exploiting the various cloud service models such as IaaS (Infrastructure as a Service), PaaS (Platform as a Service) and SaaS (Software as a Service) should be considered separately. Online social networks can be considered as SaaS cloud service. Cyber operations based on exploiting PaaS could e.g. consist of collecting and analyzing customer data, while cyber operations based on exploiting IaaS in addition would give access to entire virtual IT infrastructures of spepcific targets.

The debate around the security of cloud services is mostly centered on privacy, but national security is probably even more critical. Cloud services are frequently used even though users often do not consider then as such. For example, and online email service such as provided by hotmail.com or gmail.com is an example of SaaS. Nations states and organizations often outsource the operation of their respective DNS servers, which also can be considered to be a form of SaaS. This potentially makes these nations and organisations particularly vulnerable to a broad range of attacks should the DNS cloud service provider take part in cyber operations by other nation states.

3. Discussion and conclusion

The crucial question is whether the US and other national governments would want to take advantage of, or are already taking advantage of, collaboration with technology vendors for the purpose of conducting cyber operations. A simple mental walk-through of possible scenarios reveals that such collaboration would give considerable advantages in cyber warfare situations. For example, every running Windows operating system connects to a Microsoft server daily to check for software and configuration updates. From a technical perspective this gives Microsoft full control of every computer on the planet, as well as in space, that runs Windows and that is connected to the Internet. It would be relatively simple to identify the computers of individuals, as well as computers of a specific organization or in a specific geographical region. Remote control of each computer could be used e.g. for espionage or for sabotage, the possibilities are endless. The same type

of capabilities would be available for major microprocessor vendors Intel or AMD, where hidden hardware functions could allow remote parties to take control of computers, or telecommunication systems vendors such as Huawei or Cisco where backdoors into the systems could provide a cyber operations agent with unlimited access to network traffic. As the Snowden revelations have shown us, private corporations are willing to, or are legally obliged to collaborate with their national government, so the possibility of remote control of computers through collaboration with vendors of operating system software and hardware is realistic.

When major IT industry players become collaborators in cyber operations a natural consequence is that the market share of their products and services will be determined by global political divisions. For example, if it were the case that a US-based vendor implements a back door for the NSA in their products, and this became publicly known, then a likely result is that markets in countries that are mostly US-antagonistic would stop using these products for fear of being subject to control and surveillance.

It will necessarily have consequences when major technology vendors become collaborators in cyber operations, and this fact becomes known, either publicly or in the intelligence communities. The need for independence of nation states and for global political balance will tend to create a more fragmented global market of IT products and services. We see some of this taking place already, such as when Google during the period between 2010 and 2013 has been pressured out of the Chinese market or when Huawei is being excluded from US and Australian markets. For outsiders without access to the military intelligence communities all we see at the surface are sporadic conflicts between national governments and technology providers. These incidents are probably only the tip of the iceberg in the global politics of ICT and cyber warfare, and reflect a trend towards Balkanisation of technology.

References

107th US Congress (2001)**. U**niting and **S**trengthening **A**merica by **P**roviding **A**ppropriate **T**ools **R**equired to **I**ntercept and **O**bstruct **T**errorism Act of 2001 (PATRIOT ACT). US Public Law Pub.L. 107–56, 26 October 2001, renewed for 4 years on 26 May 2011 until 25 May 2015.

Ablon, Lillian and Libicki, Martin C. and Golay, Andrea A. (2014). *Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar*. RAND Corporation report RR610.

Appelbaum, Jacob and Horchert, Judith and Stöcker, Christian (2013). *Shopping for Spy Gear: Catalog Advertises NSA Toolbox*, Der Spiegel Online, 29 December 2013, http://www.spiegel.de/international/world/catalog-reveals-nsa-has-back-doors-for-numerous-devices-a-940994.html

Greenwald, Glenn and MacAskill, Ewen (2013). *NSA Prism program taps in to user data of Apple, Google and others*, The Guardian, 7 June 2013, http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data

McDuling, John (2012). *ASIO forced NBN to dump Huawei*. Australian Financial Review, 26 March 2012. http://www.afr.com/p/national/asio_forced_nbn_to_dump_huawei_FaglE6qWrqd5utgLpR0IdO

Menn, Joseph (2013). *Exclusive: Secret contract tied NSA and security industry pioneer*, Reuters, 20 December 2013, http://www.reuters.com/article/2013/12/20/us-usa-security-rsa-idUSBRE9BJ1C220131220

Poulse, Kevin (2013). Edward Snowden's E-Mail Provider Defied FBI Demands to Turn Over Crypto Keys, Documents Show, WIRED, 10 February 2013, http://www.wired.com/threatlevel/2013/10/lavabit_unsealed/

Roberts, Dan and Kiss, Jemima (2013). *Twitter, Facebook and more demand sweeping changes to US surveillance*, The Guardian, 9 December 2013, http://www.theguardian.com/world/2013/dec/09/nsa-surveillance-tech-companies-demand-sweeping-changes-to-us-laws

Rogers, Mike and Ruppersberger, Dutch (2012). *Investigative Report on the U.S. National Security Issues Posed by Chinese Telecommunications Companies Huawei and ZTE*. U.S. House of Representatives 112th Congress.

Sutton, Michael, and Greene, Adam, and Amini, Pedram (2007). *Fuzzing: Brute Force Vulnerability Discovery*. Addison-Wesley Professional.

Weber, Robert C. (2014). *A Letter to Our Clients About Government Access to Data*. IBM's Building a Smarter Planet Blog.. http://asmarterplanet.com/blog/2014/03/open-letter-data.html

White House (2012). *U.S. Cyber Operations Policy*, Presidential Policy Directive, PPD-20, 16 October 2012, http://www.fas.org/irp/offdocs/ppd/ppd-20.pdf

Manpower Planning and Management in Cyber Defense

Ilker Kilaz, Akif Onder and Murat Yanik
Turkish Army War College, Istanbul, Turkey
ilkerkilaz@gmail.com
akifonder@gmail.com
muratyanik76@gmail.com

Abstract: Today, cyber threats pose significant risks to governments, corporations and many other institutions as well as individuals. Tackling cyber threats requires planning and developing cyber defense strategy, enhancing Information Technology (IT) infrastructure, and more importantly managing human resources effectively. A key component in cyber defense is manpower, since it takes longer to train cyber defense personnel compared to conventional military branches such as infantry and artillery. Moreover, there is a scarcity of cyber human resources. Identifying, recruiting, training and retaining the cyber workforce necessitate a new and comprehensive approach. From the active cyber defense perspective, the cyber workforce should not have only defensive skills and capabilities, but also offensive ones such as detecting, tracing and counter-striking, mitigating and retributive counter-actions. This paper analyzes the areas in which an effective cyber manpower management needs to focus: recruiting, training and retaining cyber workforce as well as challenges and risks involved.

Keywords: manpower planning, recruiting, training, retaining, cyber workforce, cyber defense

1. Introduction

Everyday on a global scale military, energy, and many other sensitive government networks, websites and critical IT infrastructures are being attacked through cyberspace. Cyberspace is defined as "the domain characterized by the use of electronics and the electromagnetic spectrum to store, modify, and exchange data via networked systems and associated physical infrastructures" (Andress, Winterfeld and Rogers, 2011)

Cyberspace possesses many threats that stem not only from nation-states, but also from terrorist organizations, organized crime groups as well as individual cyber criminals. Nye (Nye, 2011) states that "in the physical world, governments have a near monopoly on large-scale use of force, the defender has an intimate knowledge of the terrain, and attacks end because of attrition or exhaustion. Both resources and mobility are costly. In the virtual world, actors are diverse, sometimes anonymous, physical distance is immaterial, and offense is often cheap. Because the Internet was designed for ease of use rather than security, the offense currently has the advantage over the defense." Since there are relatively low barriers to entry in cyberspace compared to conventional threats, a single hacker could wreak havoc on a nation's critical information network. Due to the proliferation of adversaries and threats in cyberspace, nations have increased their efforts to establish cyber defense organisations in order to protect their critical networks. Consistent with the power level of a country, cyber offense capabilities are being enhanced.

As cyber defense and offense capabilities are being enhanced, a new kind of war is emerging in cyberspace called "cyber war". Cyber war is "an extension of policy by actions taken in cyber space by state or nonstate actors that either constitute a serious threat to a nation's security or are conducted in response to a perceived threat against a nation's security." (Shakarian, Shakarian and Ruef, 2013)

The tools in cyberwar may seem to have a dominantly technological character in parallel with significant advances in technology. Yet solely emphasizing the technological aspect in cyber warfare will not pave the way for success. To call attention to the human aspect of cyber security, former commander of U.S. Army Cyber Command retired U.S. Army Lt. Gen. Rhett Hernandez stated that "Cyber is the number-one threat to national security, and people, not technology, will out think and out maneuver these cyber threats." (NPS Student's National, 2013) In addition, Evans and Reeder emphasize that "a critical element of a robust cybersecurity strategy is having the right people at every level to identify, build and staff the defenses and responses." (Evans and Reeder, 2010)

In this study, manpower planning and the management dimension will be examined under recruitment, training and retention. Subsequently, challenges and risks involved with the manpower aspect of cyber security will be discussed and recommendations will be put forth in the conclusion.

2. Cyber workforce

2.1 Overview

Despite the ever growing demand for cyber-related skills, a significant deficit of people with these skills remains. Moreover, the competition to identify and recruit cyber talents is fierce due to the demand from both governmental and private sector cyber institutions. According to a recent Reuters article, "Google is currently advertising 129 IT security jobs, while defense companies such as Lockheed Martin Corp and BAE Systems are looking to hire in this area." (Apps and Goh, 2013) The lack of qualifed people combined with ever increasing demand results in a staffing crunch.

Moreover, the perceptions of potential cyber warriors toward the military should be taken into consideration. According to the results of a survey (Conti and Easterly, 2010), tecnically skilled people generally expressed negative themes about the military such as: "limited creativity, limited meritocracy, technically ignorant leadership, lack of career advancement, lack of a technical career path" and so on. Thus, manpower planners need to think about how to make a cyber warfare job preferable for the candidates. To do that, the common attributes of potential cyber warriors which differ from those of traditional forces need to be closely examined.

2.2 Characteristics of cyber warfare professional

2.2.1 Education

According to studies conducted about the information security workforce "half of them have at least a bachelor's degree, with some relatively large fraction of that group having a master's degree as well. Additionally … a very small percentage with [Ph.D.] degrees." (Andress, Winterfeld and Rogers, 2011) These people also have an inclination toward learning, since they need to keep searching for new knowledge in order to stay abreast of emerging technologies; their knowledge base and skills require continuous updating.

2.2.2 Age

Age is a significant recruitment factor for the traditional forces given the rigor of military operations. But being away from the hazards of a traditional military operation, a cyber warrior may be younger or older than the average age of the conventional military personnel. Moreover the pool of candidates for the cyber workforce enlarges, since age is not a dominant factor during recruitment. (Andress, Winterfeld and Rogers, 2011)

2.2.3 Physical condition

Due to cyber warriors' differing workplace environment, the expectations for his physical condition are different than the standards expected from traditional forces. For a cyber warrior, another set of skills may be needed such as "mental abilities, creativity, technical skills, and the ability to sit in a chair for long periods of time, all the while tracking multiple activities on a series of displays, physical fitness may tend to take a back seat." (Andress, Winterfeld and Rogers, 2011)

2.3 Roles in cyber workforce

It would be misleading to consider that cyber workforce comprises of merely operators who can employ cyber tools and weapons to execute cyber attacks or defend friendly networks. Additional roles had been put forth by scholars. Franz (2011) categorizes cyber roles as follows:

- *Operators* who "plan, direct, and execute offensive and defensive activities in and through cyber space."
- *Technicians* who "provide and sustain assigned portions of cyberspace."
- *Analysts and Targeteers* who "offer intelligence support to cyber warfare operations."
- *Developers* who "design and build cyber warfare tools and weapons."

Fulp (2003) made another classification of roles in cyber workforce in terms of specilization categories:

- *Cyber tacticians* who "would focus on reducing the risk of existing fielded systems primarily through the application of appropriate safeguards (e.g., firewalls, intrusion detection, redundant configurations, data backups, etc.)."
- *Cyber strategists* who "would focus on reducing the risk of future systems primarily through the application of structured and formal system design techniques that reduce system vulnerabilities."

Fulp (2003) makes a distinction between tacticians and strategists in terms of 'existing' versus 'future' – not in the conventional military wisdom that has tactical, operational and strategic levels. He argues that cyber strategists pursue the ideal of zero vulnerability system while cyber tacticians "protect the systems that are fielded now."

The roles that are mentioned above may be sufficient only for carrying out cyber operations, but additional roles could be required to establish the integration of cyber operations with other military operations. Shakarian, Shakarian and Ruef (2013) argue that "cyber operations, in general, can aide other military operations such as intelligence gathering and information warfare." Therefore, a harmony and synchronization should be set up between military and cyber operations during the planning and execution phases. For that purpose, cyber planners may act as integrators and can perform command and control function of cyber operations.

3. Recruitment

To swiftly address the staffing problems of the cyber workforce, militaries initially sought to fill cyber positions by recruiting from their ranks. Simultaneously, they sought to attract potential cyber warriors from outside the military, including convicted hackers.

3.1 Recruiting through the ranks: Branch conversion

The arrival of cyberspace as the new domain of warfare and the proliferation of cyber threats prompted a swift reaction from governments. Since it would take a long time to build a new cyber command and cyber workforce from scratch, militaries opted to convert some of the existing communications, signal and all other relevant branches into the new cyber specialty. For example, in the U.S. Air Force, "approximately 43,000 total force enlisted airmen and 8,800 civilian personnel saw their Air Force Specialty Codes (AFSCs) alter from communications to cyberspace support on November 1, 2009." (Boland, 2010). And Israeli Parliament "approved a military restructuring plan that prods the army towards professionalisation. It urges a shift away from manpower-intensive armoured divisions in favour of the air force, intelligence collection and cyber-warfare." (Taking wing, 2013)

While branch conversion may partially help staffing the cyber workforce, converted cyber warriors may lack expertise in some areas and be "defensively focused". (Andress, Winterfeld and Rogers, 2011) Moreover, there will be a cost of training and educating these soldiers to be proficient in cyber warfare issues. Accordingly, militaries aren't expanding their cyber workforce solely from their own ranks. They are looking for the talents to fill cyber warfare positions, especially those requiring expertise, creativity and other qualifications.

3.2 Recruiting outside the military

Militaries need to reach outside and appeal to potential cyber warriors, since converting the appropriate personnel into cyber warriors insufficiently fills the cadre of cyber workforce. Presently, militaries are planning to expand their cyber workforce. For example, "U.S. Cyber Command alone plans to dramatically increase the number of personnel to 4,900 by 2015 from 900 today." (McGarry, 2013) France will be investing in cyber defense in the following years by " [bolstering] its cyber defences with a $2 billion investment in its IT infrastructure as part of the government's forthcoming military budget, according to Defence Minister Jean-Yves Le Drian**.** (…) The investment is part of a wider project to increase cyber awareness that will see personnel in the ministry's cyber division double and the number of cyber-related studies to triple." (France invests $2 billion, 2014)

By advertising and using recruiting websites and social media, militaries are trying to attract potential cyber warriors. Additionally, Conti and Easterly suggests that cyber security organizations need to reach out to the technical communities, and send "speakers to hacker conferences, competing in network warfare contests, and publishing research papers at academic conferences." (Conti and Easterly, 2010)

Reaching out the potential cyber warriors could be assisted by calling the attention of youngsters – tomorrow's cyber talents – to technical skills. In his message for Computer Science Education Week, US President Barack Obama said "Don't just play on your phone. Program it. (…) Learning these skills isn't just important for your future, it's important for our country's future. (…) If we want America to stay on the cutting edge, we need young Americans like you to master the tools and technology that will change the way we do just about everything." (Finley, 2013)

Also, militaries are considering the recruitment of hackers – sometimes even the convicted ones – and people lacking security clearances. In the United States, the Defense Advanced Research Projects Agency (DARPA) initiated a three year program in 2012 to build "experimental workshops" in 1,000 high schools. These workshops are "modeled on the growing phenomenon of "hackerspaces" — community clubhouses where hackers gather to build, invent or take things apart in their spare time." (O'Leary, 2012) Furthermore, the head of UK's Joint Cyber Reserve Unit said "convicted computer hackers could be recruited to the UK's cyber defence force if they pass security vetting." (UK cyber defence, 2013) Additionally, the International Internet System Security Certification Consortium delivered its recommendations to the US government about the stuffing crunch of cybersecurity professionals. One of the recommendations is "setting up a cyber "special forces" team designed to employ talented cyber-security workers who, whether because of personality or previous personal conduct, would not normally be able to obtain security clearances or work in a typical agency culture". (Wait, 2013)

A recent example of employing hackers and criminal organizations to carry out cyber attacks is the 2008 Russian campaign against Georgia. Cyber attacks increased the impact of Russian military campaign and "Russian criminal organizations including the RBN [Russian Business Network"] along with ""patriotic"Russian computer users – often referred to as "hacktivists"" were involved in the cyber attacks against Georgian networks. (Shakarian, Shakarian and Ruef, 2013)

4. Training

Cyber threats are growing in number and type while technological advances continue. Therefore cybersecurity professionals need to have a desire for learning, since they need to search for new knowledge in order to stay abreast of emerging technologies. Otherwise their knowledge base and skills could easily be out of date. Training the cyber workforce – either newly recruited or already existing – adds value in cyber warfare.

There' no single solution that satisfies all the training demands. ""Training today's cyber professionals requires the use of a broad range of venues to prepare these personnel to operate in a technically challenging environment," said Air Force Col. George Lamont, the director of training for the U.S Cyber Command." (Welsh, 2014) Conti and Easterly proposed such as "training with industry programs", "computer security competitions", "tuition assistance programs", free online course materials of universities, "books and lending libraries", "local guest speakers from government, academia, industry, and the hacker community", "grow[ing] professionally by leading workshops and giving talks for nearby professional groups" plus the communication abilities to "communicate technical topics to non-technical audiences." (Conti and Easterly, 2010)

An example of cooperation between cyber security organizations and universities from the U.S. is as follows: "The NSA [National Security Agency] and the DHS [Department of Homeland Security] have jointly sponsored the National Centers of Academic Excellence in Information Assurance (IA) Education (CAE/IAE) and CAE-Research (CAE-R) programs. The goal of the programs is to reduce vulnerabilities in our national information infrastructure by promoting higher education and research in IA. The programs also attempt to address the growing need for professionals with IA expertise in various disciplines." (National Centers of Academic, 2013)

Many countries expressed their strategy for developing cyber workforce in their cyber security strategies. While some of the national cyber strategies briefly talks about the importance and need to improve human capability, there are a couple of them giving detailed information about the implementation of cyber education and research activities. For example, Turkey's National Cyber Security Strategy and 2013-2014 Action Plan mandates the Council on Higher Education, the Scientific and Technological Research Council of Turkey and universities to "provide scholarship to students for postgraduate studies on cyber security, add cyber related courses to undergraduate, graduate and postgraduate curricula and establish graduate and doctoral programs on cyber security" (National Cyber Security Strategy, 2013). In addition, the UK Cyber

Security Strategy mentions "establishing programs of certified specialist training, strengthening postgraduate education, establishing a research institute in cyber security and developing cross-sector research agenda on cyber". (the UK Cyber Security Strategy, 2011)

Cyber concerns came to the attention of military circles, and military education at academies adjusted to this situation by adding cyber related courses in their curriculum and sending "teams to a variety of competitions, where they face off against each other." (Beidel, 2012)

As for the educational exercises, in the U.S. there are CyberPatriot at high school level, National Collegiate Cyber Defense Competition (NCCDC) at the college level and Cyber Defense Exercise for the U.S. Military Academies. Cyber Security Challenge was launched in the UK. Additionally, competitions are conducted in the hacker community such as Capture the Flag at DEFCON (an annual hacker convention) as well as NetWars hosted by SANS (a major commercial training company). (Andress, Winterfeld and Rogers, 2011)

Evans and Reeder puts forth a vision for cyber education and career path in their work (Figure-1). The career path has entry, practitioner, performer and expert levels. The education blocks are built accumulatively: core cyber skills being the first step and specialized training, offensive and defensive missions, support functions and the leadership block as the later steps.

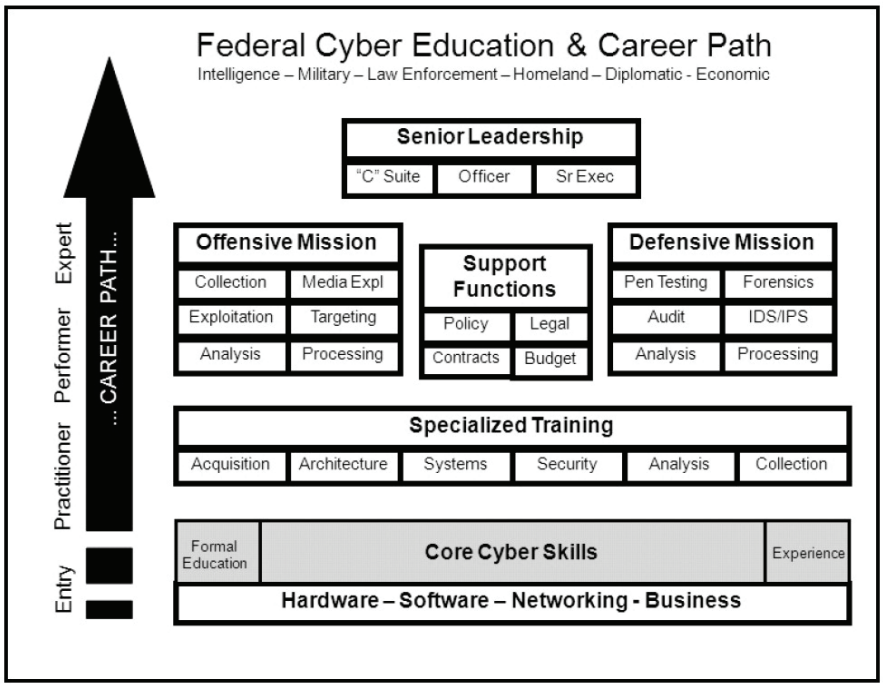

Figure-1: A vision for the learning disciplines associated with the cybersecurity workforce development (Evans and Reeder, 2011)

In another approach, core cyber skills are examined in detail and it's proposed that cyber professionals pursue the same basic education. The curriculum "should address all of the core subject matter encountered in modern information systems, and do so in a bottom-up order." (Fulp, 2003) The core subject matters – in a bottom-up order – are as follows: (Fulp, 2003)

"1) Discrete Mathematics,

2) Computer Hardware/Architecture,

3) Programming,

4) Compiler Design [compiler is a computer program that translates high-level language syntax into another programming language],

5) Operating Systems, and

6) Algorithms."

Additionally, Fulp puts forth core principles of information assurance as follows: (Fulp, 2003)

- The Reference Monitor (RM) Concept which "maintains that access control is at the heart of data protection."
- The Risk Management Equation which establishes the notion that "safeguards applied to mitigate initial risk will reduce that risk to some degree, resulting in residual risk."
- Defense-In-Depth which "dictates that practitioners of IA should not rely on any single device, technology, or security area (e.g., personnel security, physical security, etc.) when working to minimize system risks."
- The Principle of Least Privilege which states that "sensitive information should receive no more exposure to potential disclosure or modification risk than which is absolutely necessary for mission accomplishment."

5. Retention

Recruiting the appropriate people for cyber workforce could be a good start, but keeping them on board would warrant significant effort by the manpower management. There are different motives and incentives for cyber workforce. "While some cyber warriors will continue to serve because of patriotism, others to make a difference, still others because they love the mission, professional development opportunities and career progression are essential." (Conti and Easterly, 2010)

Also, the education required for a cyber warfare professional in order to stay abreast of current technology and developments in cyberspace is highly expensive if he tries to attend the courses offered by private training companies. The government cyber positions can provide requisite training to the cyber professionals without incurring the financial burdens.

In addition, the environment in which a cyber warfare professional can perform hacking legally and ethically – from a perspective of protecting the nation – can be another enticement for cyber workforce.

But the motives and incentives mentioned above may be valid provided that a viable career path, an appropriate cyber culture and a proper management style are ensured in the cyber security organization.

5.1 Career path

Evans and Reeder list four elements of strategy to address staffing crunch in cyber workforce. One of them is "assuring there is a career path as with other disciplines, like engineering or medicine, and rewarding and retaining those with high-level technical skills, both in the civilian workforce and in the uniformed services." (Evans and Reeder, 2010)

Conti and Easterly highlights the same issue by stating that "a career path that inspires others means career progression with promotions based on merit and competitive pay." (Conti and Easterly, 2010)

5.2 Cyber culture

As mentioned in the second section, the perception of the technical community about the military is generally negative. It's incumbent upon the leadership of cyber security organization to build a cyber culture appealing to potential cyber warriors. Conti and Easterly argue that "the key is to create a culture (…) to the extent possible that respects technical expertise, values diversity, and provides a viable career path from junior enlisted to General and Flag Officer." (Conti and Easterly, 2010)

Franz, on the other hand, points out the necessity of a warfighting culture in the cyber workforce. He states that cyber warriors who converted from communications and other relevant branches tend to focus "on keeping communications up and running — not on completely understanding the missions supported by each communications link or node. Consequently, true understanding of mission impact caused by losing a link or node commonly occurs only after that loss takes place and customers begin to complain." (Franz, 2011) Moreover he adds that "the culture of today's cyber warfare professionals must evolve from one that provides service to one that offers a balance of service, security, and knowledge of threats, all in the name of mission assurance." (Franz, 2011)

5.3 Managing the cyber workforce

Another important aspect of retaining the cyber workforce is how they're managed. These people are expected to think outside the box, be creative and innovative. Conti and Easterly suggest that "you don't have to overly manage them, just point in the direction you want them to go. (…) cyber warriors deserve equally tech-savvy leaders who understand and appreciate their accomplishments; who empower creative problem solving and encourage out of the box thinking; who can empower individual efforts and yet successfully focus these efforts into a cohesive team" (Conti and Easterly, 2010)

6. Challenges and risks involved

6.1 Age discrepancy

As mentioned in the second section, there can be significant age differences between the cyber workforce and the leadership. From the perspective of personnel planning and management, age does not mean much to fulfill the mission. Potentially, it would cause some discomfort between the younger personnel and seniors.

6.2 Culture clash

Conflict may occur among cyber personnel and traditional forces due to differences between set of traits expected from them. On the one hand, traditional forces are expected to demonstrate a high standard of discipline and obedience, and on the other hand cyber workforce is expected to be "creative, intelligent, [capable of] good problem solving skills, independent, and other similar terms. Unfortunately, or perhaps fortunately, depending on perspective, these types of attributes do not generally produce people that tend to follow rules well, or possess any great love of strong authority figures." (Andress, Winterfeld and Rogers, 2011)

It would cause some discomfort when some types deemed as abnormal to military culture join the cyber workforce, but all these factors must be taken into consideration when managing a cyber workforce.

6.3 Creating tomorrow's adversaries

Training and enhancing the skills of the cyber workforce in order to have a robust cyber warfare capabilities has its own drawbacks. There is a risk of investing cyber skills in the wrong person if he chooses to use his talents against his country. One may ponder the difference between the risks of training an Explosive Ordnance Disposal technician and a cyber warrior. Andress, Winterfeld and Rogers aptly articulate the issue in their work.

> *"With traditional forces, when the conflict is over, or when they have been released from their duties, the harm that can come from their training is relatively minimal. They will likely have not been allowed to depart their service with any form of advanced weaponry, although they will still be in possession of the strategic and tactical knowledge in which they were trained. (…) In cyber warfare, stripping our troops of weaponry when they leave our service may be a difficult prospect. We can certainly attempt to remove their direct access to the systems that would facilitate such attacks, but we will have likely also trained them specifically to subvert such attempts at access control. Even in the case of being able to successfully remove access to such resources, the tools with which cyber warfare is carried out are not (presently anyway) unusual by any means. More often than not, such tools are free, open source, and easily available to the public. Additionally, if we have done a proper job in our training, we should be producing people suf- ficiently skilled as to be little inconvenienced by the removal of a few specific tools." (Andress, Winterfeld and Rogers, 2011)*

The recent case of Edward Snowden is an example. In 2013, Edward Snowden, a former NSA contractor, leaked classified intelligence documents causing damage to US and UK. "Of the hundreds of documents Snowden leaked, none was more damaging than the classified document the CIA calls the "Black Budget." It's like a playbook, says [Former CIA Deputy Director Michael] Morell, revealing where the U.S. spends its money on its intelligence efforts. It would give adversaries an advantage. "They could focus their counterintelligence efforts on those places where we're being successful. And not have to worry as much about those places where we're not being successful," says Morell." (Snowden damage the worst, 2013)

7. Conclusion

The key to success in building an effective cyber force lies in attracting the candidates that have requisite skills and talents, and then recruiting, training and retaining them. There is a relatively negative perception of military among the technical community, and it makes recruiting much harder in a fierce competition for bright talents. Potential cyber warriors are expected to be intelligent, creative, innovative, and to have the ability to think outside the box. These attributes need to be taken into consideration in managing the cyber workforce.

In recruitment, militaries initially opted for converting the communications, signal and other relevant branches to cyber specialty. At the same time, they seek to fill the cyber job positions with the talents outside the military, sometimes even with convicted hackers, people lacking security clearances, criminal organizations and hacktivists.

Training the cyber warfare plays a key role in manpower management, since the cyber warrior has a desire for learning in order to stay abreast of emerging technologies. There are many venues for training: cyber security organization and academia cooperation, training with industry programs, attending courses offered by private training companies, tuition assistance programs, courses taught at military academies, educational exercises, and free online course materials of universities.

Retention of cyber warriors has three pillars: a viable career path, an appropriate cyber culture and management style. In line with these pillars, a reward system is necessary to retain cyber warriors.

On the other hand, planning and managing the cyber workforce has various drawbacks. Due to the differences between cyber and traditional forces, age does not play a dominant role for recruiting. However, the age discrepancy may create some backlash between the younger personnel and seniors. Moreover, conflict may occur among cyber personnel and traditional forces due to different set of traits expected from them. Finally there's a significant risk of investing heavily in someone who may be an opponent in the future.

Consequently, cyber domain is highly volatile, has many variables and actors as well as asymmetrical impacts. Keeping pace with cyber domain's nature warrants a new approach to manpower planning and management, since human dimension plays an important role in cyber warfare. Policy makers need to design flexible recruitment methods, ease clearance process when necessary, and establish a viable career path with effective incentives, an appropriate cyber culture and a proper management style.

References

Andress, J., Winterfeld, S., & Rogers, R. (2011) Cyber warfare techniques, tactics and tools for security practitioners, Elsevier, Waltham.

Apps, P. & Goh, B. (2013) "Cyber warrior shortage hits anti-hacker fightback" Available from: <http://www.reuters.com/article/2013/10/15/net-us-security-internet-idUSBRE99C03F20131015>. [20 December 2013].

Beidel, E. (2012) "Military Academies Look to Fill Nation's Cybersecurity Gaps" Available from: <http://www.nationaldefensemagazine.org/archive/2012/January/Pages/MilitaryAcademiesLooktoFillNation%E2%80%99sCybersecurityGaps.aspx>. [30 December 2013].

Boland, R. (2010) "Military Branch Undertakes Massive Troop Conversion", Signal, vol. 64, no. 6, pp. 49-51.

Conti, G., & Easterly, J. (2010), "Recruiting, development, and retention of cyber warriors despite an inhospitable culture". Small Wars Journal. Available from: <http://smallwarsjournal.com/printpdf/9443>. [19 January 2014].

Evans, K., & Reeder, F. (2010),"A Human Capital Crisis in Cybersecurity: Technical Proficiency Matters". Center for Strategic and International Studies (CSIS). Available from: <http://csis.org/files/publication/100720_Lewis_HumanCapital_WEB_BlkWhteVersion.pdf> [17 December 2013]

Finley, K. (2013) "Obama Says Everyone Should Learn How to Hack", Available from: <http://www.wired.com/wiredenterprise/2013/12/obama-code/>. [13 December 2013].

France invests $2 billion in cyber defences as attacks quadruple in two years, (2014) Available from: <http://www.defenceiq.com/cyber-defence/articles/france-invests-2-billion-in-cyber-defences-as-atta/>. [25 January 2014].

Franz, T. (2011), "The Cyber Warfare Professional: Realizations for Developing the Next Generation", Air & Space Power Journal, vol. 25, no. 2, pp. 87-99.

Fulp, J. D. (2003), "Training the cyber warrior", in Security education and critical infrastructures, eds C Irvine & H Armstrong, The International Federation for Information Processing, vol.125, pp. 261-273. Springer, US.

McGarry, B. (2013) "NSA Chief: What Cyberwarrior Shortage?", Available from: <http://defensetech.org/2013/10/14/nsa-chief-what-cyberwarrior-shortage/>. [15 December 2013].

National Centers of Academic Excellence, (2013), Available from: <http://www.nsa.gov/ia/academic_outreach/nat_cae/index.shtml>. [21 January 2014].

National Cyber Security Strategy and 2013-2014 Action Plan (Turkey), (2013) Available from: <http://www.ccdcoe.org/strategies/TUR_CyberSecurity.pdf>. [13 February 2014].

Nye, J.S. Jr. (2011), "Nuclear Lessons for Cyber Security?", Strategic Studies Quarterly, vol. 5, no. 4, pp. 18-38, Air University Press, Maxwell AFB, AL.

O'Leary, A. (2012) "Worries Over Defense Department Money for 'Hackerspaces'", Available from: <http://www.nytimes.com/2012/10/06/us/worries-over-defense-dept-money-for-hackerspaces.html?_r=0>. [16 December 2013]

Shakarian, P., Shakarian, J., & Ruef, A. (2013), Introduction to cyber-warfare: a multidisciplinary approach. Waltham, MA, USA, Elsevier/Syngress.

Snowden damage the worst, says ex-CIA No. 2, (2013) Available from: <http://www.cbsnews.com/news/snowden-damage-the-worst-says-ex-cia-no-2/>. [11 December 2013].

Stewart, K. (2013), NPS Student's National Cyber Professionals Association Garners High-Level Attention

Available from: <http://www.nps.edu/About/News/NPS-Students-National-Cyber-Professionals-Association-Garners-High-Level-Attention.html>. [28 Decccember 2013].

Taking wing; The Israel Defence Forces (2013), The Economist, 408 (8848), pp. 44, Available from: <http://www.economist.com/news/middle-east-and-africa/21583317-israels-armed-forces-are-shifting-emphasis-mechanised-warfare-toward-air-and>. [11 December 2013].

The UK Cyber Security Strategy, (2011) Available from: <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60961/uk-cyber-security-strategy-final.pdf>. [11 February 2014].

UK cyber defence unit 'may include convicted hackers', (2013)

Available from: <http://www.bbc.co.uk/news/technology-24613376>.[19 December 2013].

Wait, P. (2013) "Increase Cyber-Security Workforce, Government Urged**",** Available from: <http://www.informationweek.com/government/cybersecurity/increase-cyber-security-workforce-government-urged/d/d-id/1113181>. [28 December 2013]

Welsh, W. (2014), "Cyber warriors: The next generation" Available from: <http://defensesystems.com/articles/2014/01/23/next-generation-cyber-warriors.aspx>. [24 January 2014].

The Effectiveness of Online Gaming as Part of a Security Awareness Program

William Aubrey Labuschagne[1] and Mariki Eloff[2]
[1]School of Computing, University of South Africa (UNISA), Pretoria, South Africa
[2]Institute for Corporate Citizenship, University of South Africa (UNISA), Pretoria, South Africa
Aubrey.labuschagne@gmail.com
Eloffmm@unisa.ac.za

Abstract: Using cyberspace to conduct business and personal duties has become ubiquitous to an interconnected society. The use of information technology has provided humanity with a platform to evolve and contribute to the advancement of society. However duality also exists within the realm of cyberspace as shown by the expanding threats originating from cyber criminals who uses the information superhighway for nefarious purposes. Companies usually invest large amounts of money in the implementation of hardware and software controls to deter and prevent attacks on assets within these establishments. For example firewalls and anti-virus software are updated as threats evolve. In spite of these controls the weakest link in this security chain is still the human element whose actions can be considered as erratic and unpredictable thus posing a threat to the security of the organization. Security awareness programs aim to equip users of cyberspace with the necessary knowledge to identify and mitigate threats emanating from these platforms, including the Internet. Numerous security awareness frameworks exist which prescribes the required steps to design and implement an efficient and effective security awareness program. An understanding of the different steps is required to develop and customize such a program for a specific environment. Furthermore different methods which include training, newsletters and websites are used to deliver the security awareness content to the participants. The nature of these methods could be ineffective and be considered mundane and strenuous to the participants who do not always have the technical background in information technology, which, in turn could threaten the success of the implemented program. Therefore a proficient solution should be considered to attract and captivate a diverse group of employees when doing security awareness training. Moreover the effectiveness of these programs should be measured with the application of metrics defined within security awareness programs. This paper discusses the implementation and findings of a security awareness program. The aim of the security awareness program was to determine the effectiveness of using online gaming as an information security knowledge delivery method to enhance the efficacy of the participant's awareness to identify and mitigate threats encountered within cyberspace. Subsequently the paper proposes improvements to the design of the security awareness program used during the study.

Keywords: security awareness, online gaming, effectiveness, education, metrics

1. Introduction

The Internet has penetrated all aspects of daily living within society. The speed in which technology has become part of normal day to day activities created a sense of panic as users attempt to understand on how to use these new technologies. The issue of using these technologies not the only concern as cyber criminals have turned to these to prey on unsuspecting users. The arsenal available to criminals is vast and very effective against users who unknowingly would engage in actions to their own detriment as they are not aware of the threats and how to mitigate these (Kim et al. 2011). An example is the use of social networking sites. Facebook has been widely adopted and used by users to keep in contact with friends. But the Facebook platform has also been used by criminals with great success as shown by the Koobface malware which infected 400,000 and 800,000 computers in 2010 (Villeneuve, Deibert & Rohozinski 2010). In another example Labuschagne demonstrated how social media sites could also be used to profile users based on comments and posts (Labuschagne, Eloff & Veerasamy 2012). In addition users have to learn about implementing security features on a computer to protect them from network threats, for example, using a firewall. Users without the technical skills, struggle to adopt these security tools for several reasons including the user's technology adoption intention (Kumar, Mohan & Holowczak 2008). The use of a security awareness program equips computer users with knowledge to mitigate the threats that could be encountered on these platforms.

Many institutions have realized the impact of this and have started implementing awareness programs. Broadband Internet within schools in Taiwan achieved a 100% penetration in 2009. As a proactive measure by the Taiwan Ministry of Education, a security awareness program was launched to equip teachers with the necessary knowledge, whom in turn would transfer this knowledge to the scholars (Chou & Peng 2011). Computer security is defined as securing the platform from external threats and having peace of mind that the

computer system is secured (Landwehr 2001). Security awareness would entail equipping users with knowledge to identify and mitigate external threats. In another way, it is defined as being exposed to knowledge about information security related content. This newly acquired knowledge would then in turn change future behaviour. Many companies implements security awareness programs to prepare their employees for the threats originating from the digital cyber world. Several frameworks exists which provide guidance in designing, developing and deploying security awareness programs. The European Union Agency for Network and Information Security (ENISA), the SANS Security Awareness Roadmap and the National Institute of Standards and Technology (NIST) framework were some of the existing frameworks that were analyzed to determine a solution for the security awareness program used in this study.

This paper discusses the implementation of a security awareness program to determine the effectiveness of online gaming as part of such a program. The first section of the paper provides a background on the different perspectives of security awareness programs; this is followed by a discussion of the security awareness program implemented at the University of Venda in South Africa. The analysis and findings are described next and the paper concludes with recommendations to Information Security Managers who plan to deploy a security awareness program.

2. Security awareness perspectives

Security awareness programs can be designed and developed using existing frameworks. However many programs exist and the best suited framework should be selected to achieve the goals of increased security in the given domain. The ENISA framework is comprehensive and follows sequential phases which include individual steps to achieve the goal; however each phase must be completed first before continuing with the next phase. The first phase of the ENISA consists of 14 steps which includes numerous of meetings to determine the needs and identifying the goals to address the needs, selecting and recruiting a team and obtaining a budget. The second phase consists of 5 steps to execute and manage the awareness program. The last phase consists of 7 steps to evaluate and adjust the awareness program. Many of these steps require input and approval from stakeholder which potentially could increase the time to deliver the awareness program (ENISA 2010).

The SANS Security Awareness Roadmap provides an easy to interpret flow of objectives to be taken in order to implement an awareness program. This roadmap starts at no awareness program, then commences by developing an awareness program that is compliance and security metrics focussed but also promotes awareness and change resulting in long term sustainment. It provides a guide to what each objective entails and what the deliverable of each objective is. However the guide does not prescribe actions, hence additional research might be required by a novice who wants to implement a security awareness program (SANS 2010).

The NIST Awareness Framework provides enough detail to a novice to implement a security awareness program and also has been used in other studies related to security awareness programs. The NIST framework consists of four phases (Wilson & Hash 2003). The first phase entailed designing the awareness program by conducting a needs analysis. The participants of the study were students from the University of Venda. They used this security awareness program to enhance their skills as part of a community program to train people within rural areas on end user security. Another need was to determine what the current security awareness levels of the students were whom attended the awareness program. The second phase required the development of the program which included the material used during the execution of the program. The content had to address topics specific to threats end users might encounter within the cyberspace domain. The security awareness program was implemented in the third phase. This entailed identify methods to effectively deliver the material to the participants of the awareness program. The framework implementation concludes with evaluation and feedback after the program was competed.

Several studies have shown the effectiveness of security awareness programs. Eminagaoglu, Ucar and Eren (2009) implemented as security awareness program which focused on password usage of 2900 employees at a Turkish company. Three password audits were conducted over a period of a year to measure the effectiveness of the program. The results indicated a significant decrease in the use of weak passwords. Dodge, Carver and Ferguson (2007) conducted a security awareness study on phishing attacks in the United States of America. They determined that the awareness levels increased over a two year period. They developed a system that delivered phishing attacks to students and measured how many fell prey to the attack. The results showed a

decrease in successful attacks as the users become more aware of the threat. Another security awareness program was implemented at an international gold mining company which had offices in 11 countries that focussed on policies, passwords, email and Internet use, mobile devises and incident reporting (Kruger & Kearney 2006). The effectiveness of the study was measured by using multiple-choice questions provided to the participants, resulting as an indicator of awareness levels. The study also recommended that security awareness tools should consist of a broad set of questions covering the topics, using a practical system and it should be automated. These recommendations were considered during the design and development of the online game used within this study. A study conducted on the ever changing information security domain highlights the importance of the continual delivery of security awareness programs as a mechanism to equip computer users with knowledge to deal with cyber threats (Dlamini, Eloff & Eloff 2009). In other words, security awareness programs should not be a once off event as threats evolve with the rapid change in technology. This is supported by Rezgui and Marks who explored factors that effects security awareness (Rezgui & Marks 2008). They found that working environments does play a role and that iterated training should be regular to be effective.

Security awareness content is delivered using various methods including, but not limited to, posters, classroom-style training, websites and newsletters. Subsequently the effectiveness of security awareness programs also needs to be measured as this mechanism could improve future security awareness programs. A study conducted by Khan (2011) on the effectiveness of the different security awareness methods listed, group discussions and educational presentations are the most effective. Other methods, such as email messaging, newsletters, video games, computer-based training (CBT) and posters were not as effective when looking at knowledge gained, attitude to change, subjective norms, change in behaviour and a component of intention. However, video games have been used in security training as they draw the attention of the users and allow the users to implement knowledge acquired within a given scenario (Cone et al. 2007).

Most end users access services on the Internet using a web browser. The majority of users either access their email, social networking site accounts or visiting web sites. Subsequently types of attacks were associated with each vector also known as a vulnerability. The content of the security awareness program were designed to address each of the potential threats. The identification of the topics resulted from the development an attack tree by the authors to determine the different vectors that could be used with malicious intent against unsuspecting end users. The graphical representations of the different attack vectors are depicted in

Figure 1.

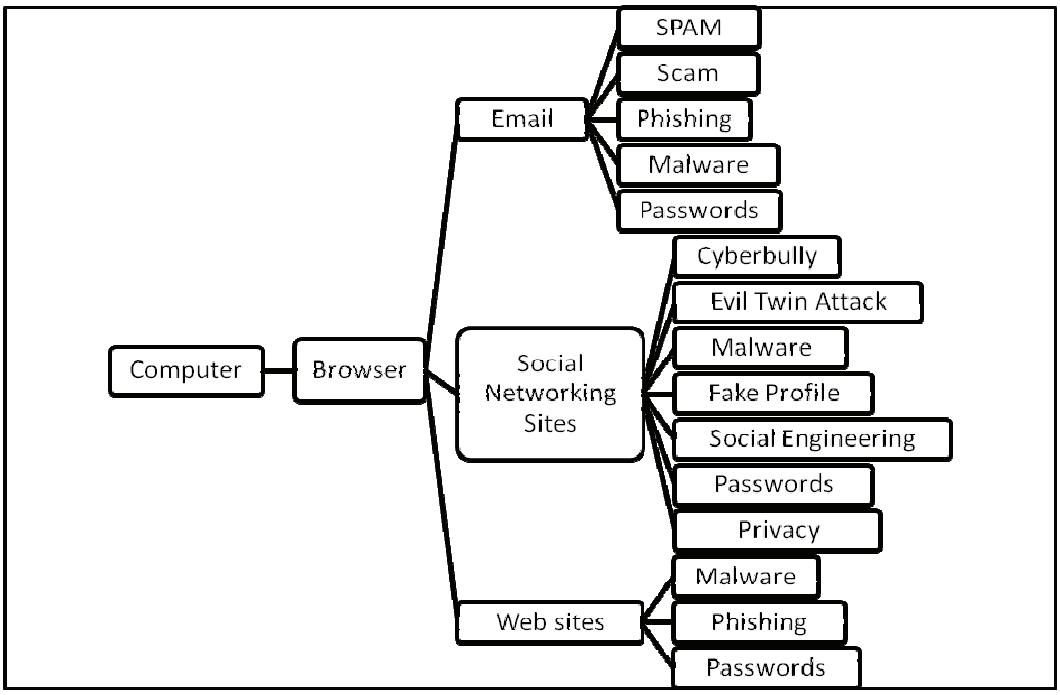

Figure 1: End user attack tree

The security awareness program discussed in this paper covered the following topics which subsumes the potential threats identified by the attack tree: Web Browsers, Passwords, Social Networking Sites, Cyber Bullying, Malware and Phishing.

The threats listed by the authors align with the threat landscape described by Veerasamy and Taute (2009) who conducted research in identifying what attack strategies and techniques were used against national, commercial, governmental and individual entities. The next section describes the implementation of the security awareness program.

3. Method

The participants were from 3rd year Computer Science class at the University of Venda, Thohoyandou, South Africa, but also formed part of rural development plan to educate people from the area about the different security threats encountered within the digital realm. A class of 40 participants attended the security awareness program. All participants were in the same venue for the day session. The security awareness program was initiated with the first survey (Pre Assessment) which also resulted in determining the awareness level baseline of the participants. This was followed by a training session which covered security topics identified during the analysis. The training sessions did not exceed 15 minutes per topic as it was a precautionary measure against mental fatigue (Wilson & Korn 2007). A second survey (Post Assessment 1) followed the completion of the training session. The objective of the second survey was to determine if the training session had an impact on the overall security awareness levels of the group. The program continued with the participants playing a social networking game online. The game was designed and developed with gamification concepts in an attempt to improve the retention of the content from the topics discussed during the training session. Some of the design concepts included but was not limited to a leaderboard indicating who has attained the most points, a progressbar to provide a graphical feedback on how far the user was in completing the game and a timeline to provide all users with a view of other events that occurred within the game. A prize was handed over to the winner of the online game and the security awareness program was concluded with the final survey (Post Assessment 2). The objective of the last assessment was to identify if the online game had an effect on the awareness levels of the participants.

Data was collected during the completing of the surveys as well as playing the online game. The surveys were accessed online by the participants and consisted of seven sections with five questions per section which totalled to 35 questions per survey. Each survey was designed to address the same objective across the different surveys for example, to determine if the participant understood the concept of weak passwords. Access to the surveys where controlled with a unique token number and was individually issued to each participant.

The online game was hosted within Facebook and the objective required from the users was to acquire points by answering security awareness related questions which correlated with the topics used during the awareness program. Random events occurred within the game which could deduct points from the user's accumulated points. An example of such events was a virus infection or hard drive failure. Users could prevent loosing points by obtaining an item which, once in possession, would counteract the event, for example, the virus infection would be negated if the user has an anti-virus item. Users could buy items by using points accumulated hence losing some points in order to prevent a substantial loss of points when an event occur.

4. Findings

All the survey data was analyzed for each session and the results are depicted in **Figure 2**. Each item in the survey had a correct answer. Each participant's survey results were programmatically calculated to determine what topics each participant understood and what individual topics needs to be revised. This was seen as the awareness level of the individual. The group's awareness level was determined by averaging all the results of the group for each survey.

The first assessment which is used as the baseline shows the groups awareness levels at about 21 correct answers out of a possible 35. The participants then attended the training session which after the second assessment resulted in about the same number of correct answers as before the training session. Next the group participated in playing the online game that focused on the security awareness topics. The results of the third assessment after playing the game, show a considerable decline in correct answers. It was expected that

the number of correct answers should have increased which would support the notion that playing the online game should increase the retention of the knowledge on the security awareness topics. Astoundingly the opposite happened and the awareness levels of the group decreased.

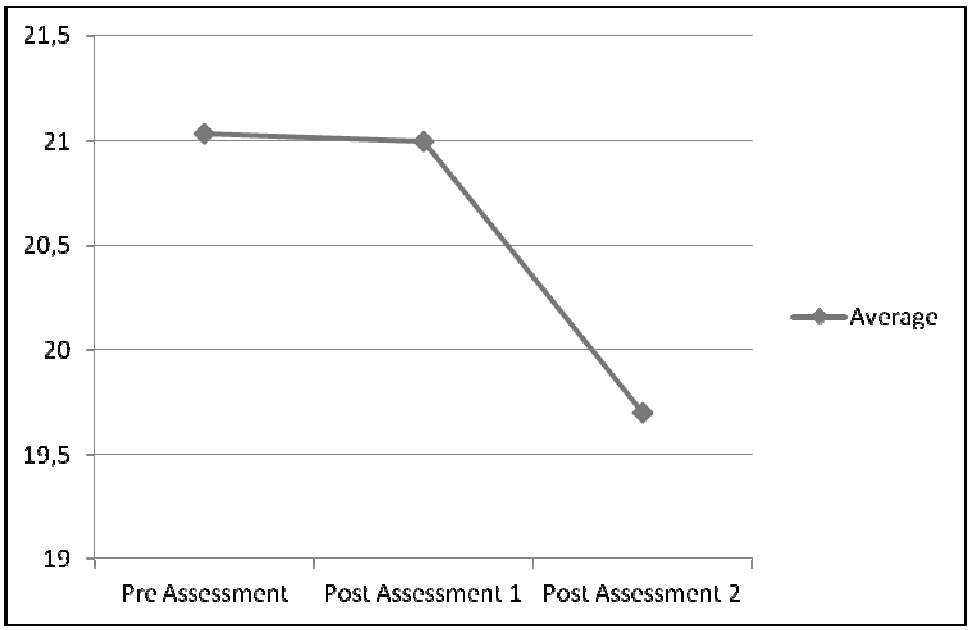

Figure 2: Session averages

Next the online game data was analysed to determine if the results of the assessments are accurate in showing the negative effect of the online game on the group awareness levels. **Figure 3** shows the number of responses received during the playing of the game. Noticeable is the increase in responses as the deadline approached for the game play, which is mainly attributed by the prize which was handed to the participant with the most points.

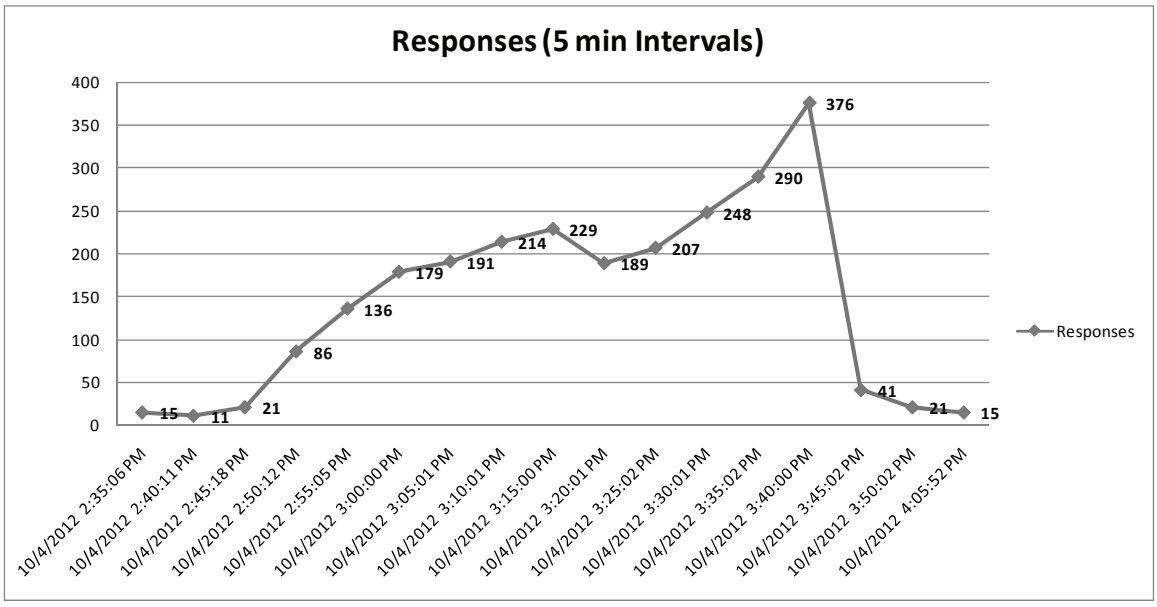

Figure 3: Average response time

The responses were classified to indicate how many were correct and incorrect (as depicted in **Figure 4**). Surprisingly the number of correct answers increased as the number of responses increased. This finding shows the participants knowledge on the awareness topics increased with time. Hence if the participants did not learn during the game play then the number of correct answers should have decreased and the number of incorrect answers should have increased. This could be due to the spacing effect as the recall of knowledge will decrease as time elapses. But the more the participants are exposed to the topic the longer they tend to remember it (De la Rouviere 2012).

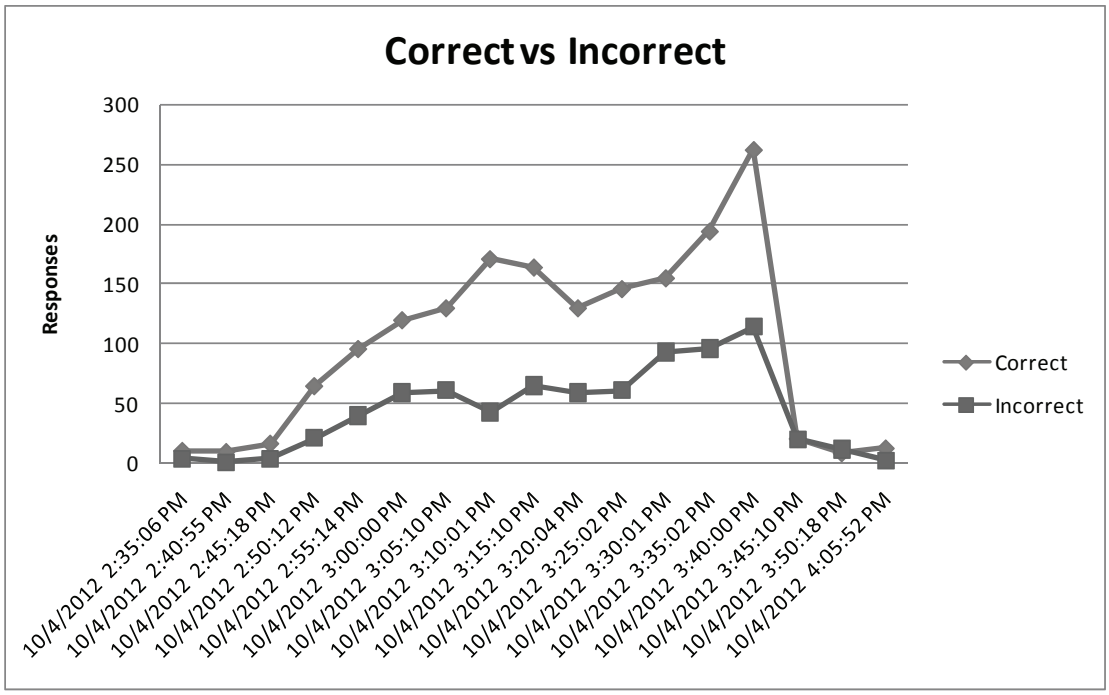

Figure 4: Response classification

One of the metrics captured by the online game was the time it took for participants to read the question and then answer with the response. As the participants were in the same venue, the network latency affected all in the group the same way, therefore the response time calculations are consistent. An interesting observation shows that correct answers took a shorter time to answer than incorrect answers, even when more responses were submitted by the participants (See **Figure 5**). This also supports the notion that knowledge increased while playing the online game. Participants who did not know the answer had to consider the different options thus increasing the time to answer the questions. Participants, who had the knowledge, instinctively answered the questions quicker, hence a faster response time.

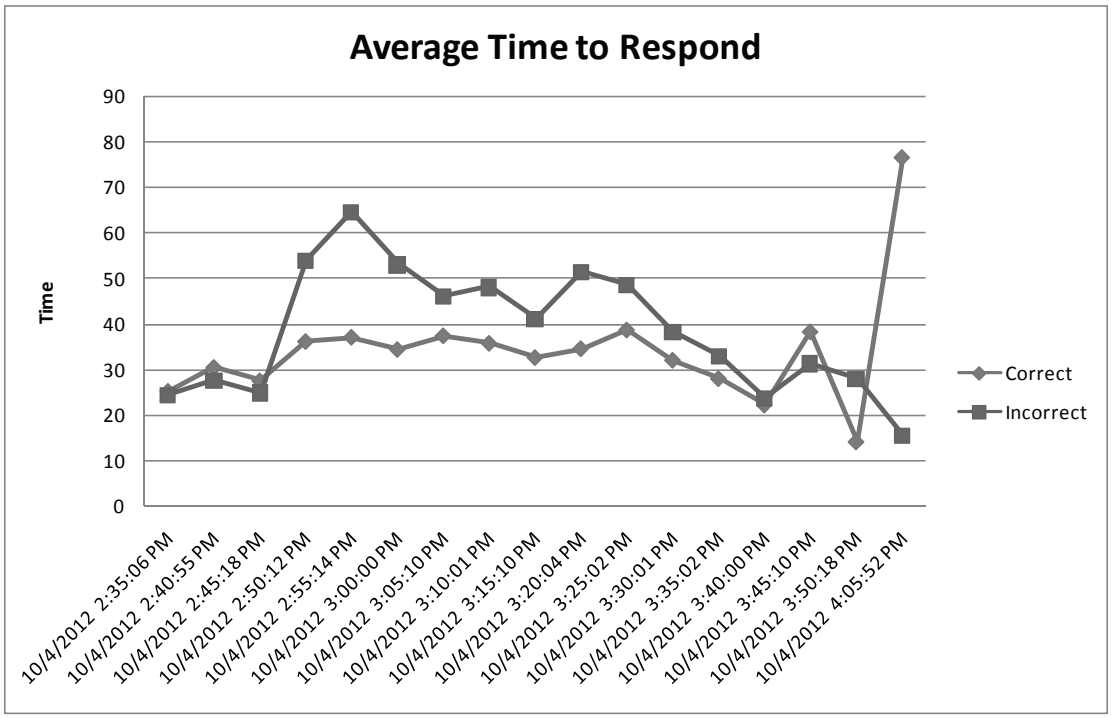

Figure 5: Average response time

Next the different topics answered in the online game were compared against the results from the second assessment (Post Assessment 1) and the last assessment (Post Assessment 2). The findings are depicted in

Figure 6. The online game results show that all the topics except for Phishing and SPAM/Scam outperformed the individual topic results in the other assessments (Post Assessment 1 and 2). Also, if the online game had a positive effect on the awareness levels then the results of the last assessment (Post Assessment 2) should have improved against the results of the second assessment (Post Assessment 1). This was the finding except for Phishing, SPAM/Scam and malware however this is inconclusive in proving the positive effects of online gaming within this study however as discussed earlier the online game did indicate an increase in knowledge.

Alternatively the effect of the extrinsic motivation should be considered. The participants were competing for a prize as an incentive to partake in the study which was given after the completion of the online game. This would imply that the motivation for the students to partake in the study declined once the prize was handed over to the winner of the online game hence negatively affecting the results of the last assessment. This finding is aligned with results by Deci who examined the effect of extrinsic rewards on intrinsic motivation (Deci 1971). Subsequently the effects of the online game cannot be inferred due to the skewed results recorded during the last assessment.

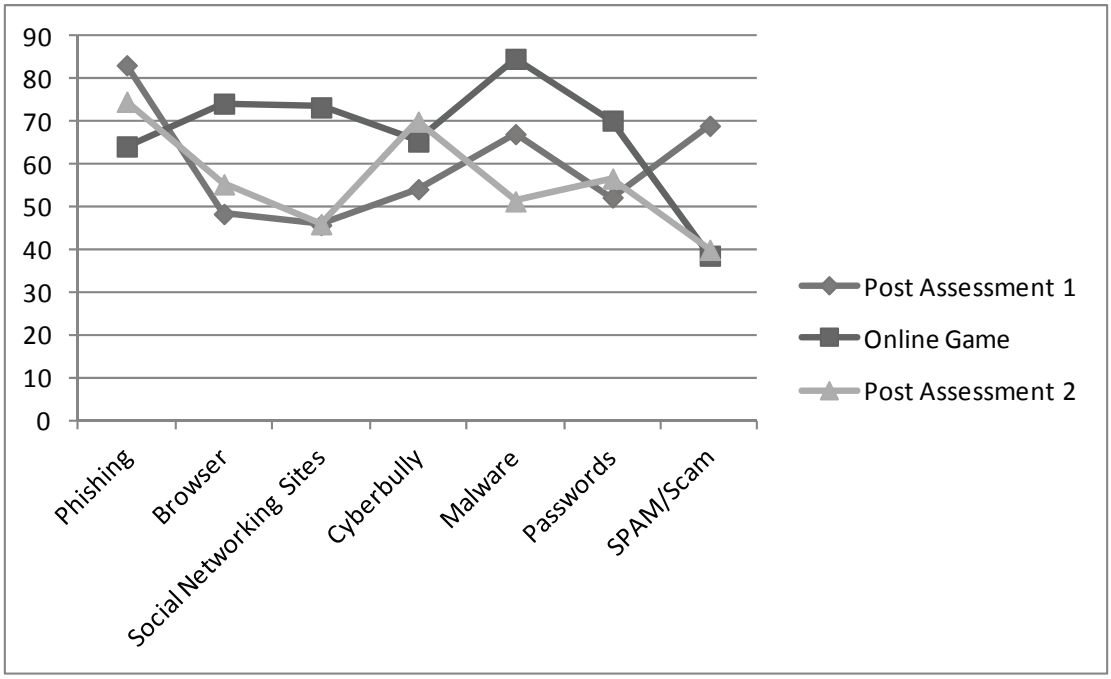

Figure 6: Comparing online game with assessments

5. Conclusion

Technology has become part of everyday living and society has embraced it as these digital platforms, which include mobile devices and the Internet, improves living conditions. For example many errands can be completed by merely logging into an Internet connected devise and accessing services for example paying bills without leaving the comfort of their home. The duality of technology has become apparent as cyber criminals have seized the opportunity to use these platforms for nefarious purposes. Many users do not have the knowledge to identify these threats and mitigate it before harm is done. Security awareness programs are aimed at users who frequently use these digital platforms and to equip them with the appropriate knowledge to mitigate threats encountered within the cyberspace domain. Due to the nature of the computer domain, many users are not interested in the technical aspects as it is deemed complex and not interesting enough to engage the attention of the end user who is not technically inclined. This is detrimental to security awareness programs. These programs are delivered using various methods which include posters, training, presentations and websites. Games have been widely used to teach users about various topics as it is deemed fun. The use of games also provides a good indication if the user can implement new acquired knowledge within an environment. The effectiveness of games within security awareness programs was pursued in this study. The participant's security awareness levels were measured by using questionnaires that focussed on the different topics identified for the awareness program. The desired outcome of having an increase in the group awareness levels after the completion of the game play was not achieved. An investigation revealed that extrinsic rewards could have affected the intrinsic motivation which subsequently caused the participants to

lose interest after game session. This subsequently meant the last questionnaire results which were critical in the study were affected. Thus it is the opinion of the authors that the final assessment (questionnaire) skews the results and cannot conclusively demonstrate that the gaming session can improve the security awareness levels of the participants. The study should be conducted again in future and only hand over the reward after the completion of the full program. However, analysis of the gaming session provided numerous findings that indicate learning has occurred. Participants substantially increased the number of responses as the session concluded, noticeably the correct responses increased as well. If the participants did not learn during the gaming session, then the incorrect responses would have increased. Another finding was the time for correct responses was shorter and consistent than incorrect responses which took longer. These findings provide feedback on improving future security awareness programs.

References

Chou, C. & Peng, H. 2011, "Promoting awareness of Internet safety in Taiwan in-service teacher education: A ten-year experience", The Internet and Higher Education, vol. 14, no. 1, pp. 44-53.

Cone, B.D., Irvine, C.E., Thompson, M.F. & Nguyen, T.D. 2007, "A video game for cyber security training and awareness", Computers & Security, vol. 26, no. 1, pp. 63-72.

De la Rouviere, N. 2012, 30 Jan 2012-last update, How To: Learn Better by Learning Less. [Homepage of MIH Media Lab], [Online]. Available: http://ml.sun.ac.za/2012/01/30/how-to-learn-better-by-learning-less/ [2013, 09/23].

Deci, E.L. 1971, "Effects of externally mediated rewards on intrinsic motivation.", Journal of personality and social psychology, vol. 18, no. 1, pp. 105.

Dlamini, M.T., Eloff, J.H.P. & Eloff, M.M. 2009, "Information security: The moving target", Computers & Security, vol. 28, no. 3–4, pp. 189-198.

Dodge, R.C. 2007, "Phishing for user security awareness", Computers & Security, vol. 26, no. 1, pp. 73-80.

Eminagaoglu, M., UÃ§ar, E. & Eren, S. 2009, "The positive outcomes of information security awareness training in companies-A case study", Information Security Technical Report, vol. 14, no. 4, pp. 223-229.

ENISA 2010, The new users' guide: How to raise information security awareness., European Network and Information Security Agency (ENISA).

Khan, B., Alghathbar, K.S., Nabi, S.I. & Khan, M.K. 2011, "Effectiveness of information security awareness methods based on psychological theories", African Journal of Business Management, vol. 5, no. 26, pp. 10862-10868.

Kim, W., Jeong, O., Kim, C. & So, J. 2011, "The dark side of the Internet: Attacks, costs and responses", Information Systems, vol. 36, no. 3, pp. 675-705.

Kruger, H.A. & Kearney, W.D. 2006, "A prototype for assessing information security awareness", Computers & Security, vol. 25, no. 4, pp. 289-296.

Kumar, N., Mohan, K. & Holowczak, R. 2008, "Locking the door but leaving the computer vulnerable: Factors inhibiting home users' adoption of software firewalls", Decision Support Systems, vol. 46, no. 1, pp. 254-264.

Labuschagne, W.A., Eloff, M.M. & Veerasamy, N. 2012, "The dark side of Web 2.0", IFIP Advances in Information and Communication Technology, vol. 386/2012, no. ICT Critical Infrastructure and Society, pp. 237-249.

Landwehr, C.E. 2001, "Computer security", International Journal of Information Security, vol. 1, no. 1, pp. 3-13.

Rezgui, Y. & Marks, A. 2008, "Information security awareness in higher education: An exploratory study", Computers & Security, vol. 27, no. 7-8, pp. 241-253.

SANS 2010, Security Awareness Roadmap, SANS Institute.

Veerasamy, N. & Taute, B. 2009, "Introduction to emerging threats and vulnerabilities to create user awareness", Information Security South Africa (ISSA), Johannesburg, South Africa.

Villeneuve, N., Deibert, R. & Rohozinski, R. 2010, Koobface: Inside a crimeware network, Munk School of Global Affairs.

Wilson, K. & Korn, J.H. 2007, "Attention during lectures: Beyond ten minutes", Teaching of Psychology, vol. 34, no. 2, pp. 85-89.

Wilson, M. & Hash, J. 2003, Building an Information Technology Security Awareness and Training Program, National Institute of Standards and Technology, Gaithersburg.

Cyberconflict and Theoretical Paradigms: Current Trends and Future Challenges in the Literature

Andrew Liaropoulos
University of Piraeus, School of Economics, Business and International Studies, Department of International and European Studies, Piraeus, Greece
aliarop@unipi.gr
andrewliaropoulos@gmail.com

Abstract: The purpose of this paper is to critically review the literature on cyberconflict and identify conceptual and empirical gaps that can help guide future research in this specific field of security. The argument of this paper is that cyberconflict is a phenomenon with both technological and socio-political aspects. Therefore cyberconflict encompasses both cyber-attacks in the critical information infrastructure, but also the use of social media networks in mobilizing for collective action during conflicts. Based on this thesis we identify two paradigms in the relevant literature: the (r)evolutionary paradigm that questions the degree in which cyberspace has transformed the nature of war and the social wave paradigm that reviews the impact of social media tools in socio-political conflicts. The criterion for categorizing the various approaches in the literature into these two paradigms, is the way cyberspace, power and conflict interrelate with each other. This categorization is useful not only in identifying the differences between each paradigm, but rather in highlighting the conceptual and empirical shortcomings in the present literature.

Keywords: cyberconflicts, cyber security, (r) evolutionary paradigm, social wave paradigm, cyber power

1. Introduction

Cyber security is without a doubt one of the major security challenges of our time. Cyberconflicts have attracted the attention of scholars from various disciplines. Technical and operational issues regarding the conduct of cyber-attacks, legal aspects concerning the need of international law to regulate cyberconflicts, ethical questions relating to the just way of conducting cyber-operations, the use of social media networks to accelerate policy changes, as well as concerns about the quest for power and security in the cyber domain, are constantly being debated. As a result, over the past two decades, there is a growing body of literature that offers intellectual depth and covers adequately several features of cyberconflict. Nevertheless, there are conceptual and theoretical shortcomings in the present literature.

The purpose of this paper is twofold. On the one hand, to frame the debate on cyberconflicts, by categorizing the main approaches into working paradigms, and on the other hand, to identify challenges for future research in the field. Therefore, the first part of the paper will define the concept of cyberconflict and present the two paradigms, labelled for the purposes of our analysis as the (r)evolutionary paradigm, and the social wave paradigm. The principal value in such an effort is to organize the various arguments and to sort out the areas of agreement and disagreement. The term paradigm is used here to describe a theoretical framework, a set of hypotheses that will serve as an organizing principle for our analysis. After critically reviewing the literature through the prism of the two paradigms, the second part of the paper will define theory gaps and address topics that deserve closer attention, like the ambiguous nature of cyberspace and the concept of cyber power.

2. Cyberconflicts and theoretical paradigms

Over the past two decades, there has been a breathtaking evolution of cyberspace that has impacted upon almost every aspect of our lives. Due to the continuing development of Information and Communication Technologies (ICTs) and the decentralized architecture of Internet, cyberspace has experienced a rapid expansion from a small network designed primarily to serve the scientific community, to a global network, that manages billions of users (Karatzogianni 2006, Cavelty et.al 2007, Costigan & Perry 2012, Choucri 2012). Cyberspace obscures the identity and location of its users and increases the speed, volume and range of communications. Cyberspace has undoubtedly changed the ways that individuals interact with each other, how companies deliver services, how people are governed and even the way societies perceive security, power, sovereignty and international order (Kramer et.al 2009, Betz & Stevens 2011, Liaropoulos 2011, Liaropoulos 2013). The cyber realm offers new opportunities for cooperation, competition and conflict.

Regarding the latter, state and non-state actors choose cyberspace as a new battlefield, where socio-political, economic and military conflicts are increasingly being carried out. Although scholars vary on the way they approach and define cyberspace, they all agree that cyberspace consists of three layers. The first one is the physical layer that consists of electrical energy, integrated circuits, communications infrastructure, fiber optics, wires, transmitters and receivers. The second layer is the software, meaning the computer programs that process information. The last and least concrete layer is that of data (Tabansky 2011: 77). This layer, labeled by Martin Libicki as semantic contains the data - information that is meaningful to human beings. It is in this layer that a different version of reality can be presented (Libicki 2007: 9). The semantic layer can be understood as the cognitive function of cyberspace, where information can shape perceptions and social behavior (Langø 2013: 10). Cyberconflict has to be approached in all three layers.

Attacks in cyberspace come in many different forms, and their destructive potential is limited only by the creativity and skill of the attackers behind them. The digital battle-space is comprised of many components. These would include the following: web servers, information systems, communication links, network equipment, and the computers in businesses and homes (Liaropoulos 2011: 37). The terrain also encompasses information systems like the electrical grids, telecommunication systems, but also social media platforms that play a role in organizing civil unrest or serve as a propaganda tool. Anyone attempting to untangle the complexities of any socio-political conflict cannot afford to ignore the importance of social media intelligence (Rovner 2013, Liaropoulos 2013a).

Cyberspace, in common with the physical domains of land, sea and air, has proven to be no stranger to conflict and great power antagonism. There is no doubt that critical national infrastructures that depend upon computer networks have become increasingly vulnerable to cyber-attacks. The global economy is becoming ever more susceptible to cyber-crime and espionage. Yet, criminal networks, terrorist groups and hackers, are only part of the picture. As cases such as the cyber-attacks on Estonia in 2007 and Georgia in 2008 - not to mention the advent of Ghostnet and Stuxnet - demonstrate, armed forces are also capable of launching hard-to-detect cyber-attacks across national borders (Liaropoulos 2014). Another aspect of cyberconflict involves the role of social media in mobilizing for collective action and political protest. Social revolutions that took place in many countries over the past years, demonstrate that bloggers, web activists and average citizens, acting as online citizen-journalists, broadcasted amateur videos, audio interviews, tweets and emails, producing a near real-time flow of information that reached millions of users around the world. The so-called Twitter and Facebook Revolutions vividly illustrate the utility of these tools regarding the organization, mobilization and recruitment through cyberspace (Hussain & Howard 2013, Liaropoulos 2013a).

Therefore, for the purpose of the present analysis we adopt a rather holistic approach regarding the concept of cyberconflict. In particular, we believe that the term cyberconflict should involve not only the attacks on the critical information infrastructure of a society, but also the way social media networks are utilized in socio-political conflicts. The reason is that such an approach offers a synthesis of both the technological and the socio-political aspects of cyberspace as a medium for conflict (Karatzogianni 2006, Karatzogianni ed. 2009). It is hard to imagine nowadays a socio-political confrontation that does not involve some element of cyber-activity. Cyberconflict involves all actions that take place in cyberspace or are materialized through cyberspace, in both wartime and peacetime, and encompasses cyber activities such as sabotage, espionage and psychological operations.

The relevant literature has tried from the early days to investigate whether the cyber domain is radically transforming the concept of security and the nature of conflict. Opinions differ, since scholars approach the topic from various angles and shed light on several aspects that relate to cyber security. The aim here is to categorize the most influential approaches into two working groups that sufficiently address the debate about the nature and the future of conflicts in cyberspace. The appropriate way to use these paradigms is to treat them as arguments. They are not valid as prescriptions; rather, they describe categories of thinking. Therefore, they instruct the scholar in what to think about, but not what to conclude. The categorization of the literature into these groups is based on the way they approach the interrelation between cyberspace, power and the nature of conflict.

2.1 The (r)evolutionary paradigm

Scholars in the (r)evolutionary paradigm discuss the relationship between cyberspace and security and explore whether ICTs are truly transforming the nature of war. Based on how they approach these issues they can be characterized as revolutionists or evolutionists.

The revolutionary paradigm views cyberspace as a new domain, where the nature of conflict and security are, or will soon be, radically transformed. Scholars that belong in this paradigm place great emphasis on the ICTs and their impact on military doctrines, organizational principles and on the war-fighting model in general. Their thesis is that technology has radically shaped warfare in the past and ICTs will trigger a revolution in the way modern societies conduct war. Throughout history, technology has been a catalyst that has brought about a series of radical transitions from the Blitzkrieg doctrine to Network Centric Warfare and from the artillery revolution to the stealth revolution (Langø 2013a: 9-15). At the forefront of this paradigm are scholars that have written extensively on the Revolution in Military Affairs theory and the concepts of Information Warfare in the 1990s (Libicki 1995). Proponents of this paradigm refer to a digital battle-space, which is not dominated by mass troops and conventional armour, but cyber-bytes and cyber-warriors (Bunker 1996, Arquilla & Ronfeldt 1997, Rattray 2001). Cyberspace is a new domain of conflict in which both states and non-state actors can launch cyber-attacks. Cyberconflict can be contained within cyberspace or can also take place in the physical realm, causing economic and physical damage. Cyberspace is structured as an open network that allows for anonymity as well as deniability and thereby makes it difficult to attribute cyber-attacks. Part of the scholarship that belongs in the revolutionary paradigm stresses the destructive potential of cyber-attacks against the critical infrastructure (electrical power, pipelines, airlines, banking, etc.) and draws the parallel with Pearl Harbor and 9/11 (Schwatz 2007, Clark & Levin 2009).

Policymakers seem to accredit such views (Langø 2013a: 16). The U.S Secretary of State John F. Kerry described cyber-attacks as 'the 21st century nuclear equivalent' (Kerry 2013) and former U.S Deputy Secretary of Defense William J. Lynn III stressed the extensive vulnerabilities in cyberspace and the lack of credible deterrence (Lynn 2010). Richard Clarke, the former U.S National Coordinator for Security, Infrastructure Protection and Counterterrorism, pointed out the possibility of an electronic Pearl Harbor. According to Clarke, cyber-attacks targeted at hijacking critical infrastructure are particularly alarming (Clarke & Knake 2010).

Two classical examples that revolutionists use to support their arguments about the way in which cyberspace has transformed warfare, are the cyber-attacks that took place in Estonia in 2007 and the Stuxnet worm. The cyber-attacks in Estonia were the first that were possibly directed as a coercive instrument in a political conflict between Russia and Estonia. In April 2007, the Estonian government's decision to move a Soviet-era war memorial, the Bronze Soldier, triggered a conflict, in the form of a three-week wave of distributed denial-of-service (DDOS) attacks that crippled the country's information technology infrastructure. The cyber-attacks temporarily disrupted the Estonian communications networks, by targeting the government, newspapers, mobile phones, emergency response systems and banks. The target included the Estonian presidency, its parliament and many government ministries (Blank 2008: 227-247). The Estonian case also vividly highlighted the issue of attribution (Tsarougias 2012). Although the cyber-attacks cannot be attributed to a specific actor, it is widely believed in Estonia that Moscow was behind these attacks.

Stuxnet is malicious software that was designed specifically to strike the Iranian nuclear facility at Natanz. It was designed to sabotage gas centrifuges - the machines that can enrich uranium for use in both nuclear reactors and nuclear bombs. It spread via Microsoft Windows and targeted Siemens industrial software. The value of the Stuxnet lays not so much on its technical characteristics, but on the political and strategic context, within which it operated. Stuxnet offered a better and risk-averse alternative to a conventional attack (Farwell & Rohozinski 2011: 23-40).

This brief overview demonstrates that the revolutionary paradigm is popular not only in the academic circles, but also among policy-makers. After all, the trend in drafting, implementing and evaluating international and national cyber security strategies is only growing. Identifying vulnerabilities and achieving a higher level of protection against rapidly evolving cyber threats is of course welcome. The revolutionary paradigm is quite influential and has to a certain extent defined the debate about the interrelation between ICTs, cyberspace and the nature of war. Nevertheless, the revolutionary paradigm has received some harsh criticism over the last years.

According to some authors, a significant shortcoming in the revolutionary paradigm is that there is no common understanding of the terminology that is used. Scholars, but also policymakers, use the term cyberwar as a catch-all phrase that includes a variety of activities, ranging from attacks on critical infrastructure and espionage to sabotage and propaganda. The effects of such operations vary, depending on whether they are violent or not, whether they attack civilian or military targets and whether they spill over into the physical realm or not. Adding to that, a second limitation is that much of what is analysed in the revolutionary paradigm is largely speculative. The cyberconflicts in Estonia and Georgia or even Stuxnet do not offer efficient empirical evidence to value the true scale of transformation. In contrast to conventional wisdom, the analysis of the Stuxnet case reveals that the worm was not that effective in causing a major setback to Iran's uranium-enrichment programme and on the contrary, may have been of net benefit to Tehran (Barzashka 2013). In large, there is a disagreement over the nature and the severity of the threats in cyberspace.

Based on these criticisms, we can identify another set of arguments in the relevant literature that is labelled as the evolutionary paradigm. Scholars in this paradigm, are quite sceptic about the true effect of the ICTs on the nature of war and security. Proponents of the evolutionary paradigm argue that technological developments in general (and ICTs in particular) are important in defining warfare, but as always they consist only one part of a complex equation that defines strategy and warfare (Knox & Murray 2001, Gray 2001, Williams 2001). The evolutionists adopt a holistic understanding of war that places the technological advantages within a broader strategic context and criticize the hype over cyberwar (Cavelty 2007, Lawson 2011). Very often, policy-makers, produce policy recommendations based on doomsday scenarios and misplaced rhetoric (Brito & Watkins 2011).

Supporters of this Clausewitzian approach argue that the uncritical emphasis on a single dimension (ICTs) is bound to lead to strategic failure (Lonsdale 2004). Their scepticism derives from recent and past experience. In the past, air power or blitzkrieg, produced a short-lived transformation in the way military operations were conducted, but in the long run, they were just evolutionary moments in the history of war. Likewise cyberwar or network centric operations should not be seen as silver bullets, but rather as one more case in the long list of evolutionary steps in the transforming character and not nature of war (Gray 2013). Echoing Martin Van Creveld, the underlying logic of technology is linear, whereas the underlying logic of war is paradoxical (Van Creveld 1991: 319). Since technology and war operate on a different logic, it is counterproductive to perceive cyberwar as simply an extension of ICTs. Based on Clausewitz's definition of war as violent, instrumental and attributable to one side as an action taken for a political goal, evolutionists scholars, question whether - worms, denial of services (DOS) and trojan horses that are non-violent and non-attributable - constitute real war (Rid 2012, Gartzke 2013).

To sum up, for evolutionists the idea of war still stands for violence, destruction and suffering. Classifying disruptions of websites or economic espionage as war is misleading and only adding to the conceptual confusion about the true nature of threats in cyberspace. According to Thomas Rid, sabotage, espionage and subversion are more suitable terms to describe the present typology of cyber-attacks (Rid 2012). Cyber-attacks represent an evolutionary change in the way war is conducted and add one more tool in the toolkit of warfare.

The (r)evolutionary paradigm is useful in highlighting the connection between ICTs, war and security in cyberspace. The revolutionary scholars have enriched the discussion about the impact of ICTs and their work has influenced policy-makers and the current thinking on cyber security. Evolutionist scholars on the other hand have been sceptical about the true impact of ICTs and have raised legitimate questions about the concept of cyberwar. The evolutionists also correctly point out that war is a continuation of politics, a form of political violence, where cyberspace and cyber-attacks are the means and techniques of conducting war (Langø 2013a: 23). Cyberwar as any other type of war needs a political context to occur and cyber-attacks are just a means to a political end. Therefore it is important to give cyberwar a political context. To achieve this, it is worth taking a different and broader approach to cyberwar, one that places emphasis on the socio-political aspects of cyber conflicts.

2.2 The social wave paradigm

The second paradigm places emphasis on the socio-political aspects of warfare and perceives cyberconflict as the outcome of a great social wave change. The origins of this paradigm can be traced back in the early 1990s, when Alvin and Heidi Toffler argued about the relationship between the three waves of civilization (the

agrarian, the industrial and the information wave) and warfare, and the shift from one wave to the other. According to them, the way a society makes war reflects the way it makes wealth. Starting with the very invention of agriculture, every revolution in the system for creating wealth triggered a corresponding revolution in the system for making war. The present economy is based on knowledge and this remarkable change is bringing with it a parallel change in warfare (Toffler 1980, Toffler & Toffler 1993). Proponents of this paradigm argue that microelectronics, data processing and communications technologies are creating a third wave society, where information technology replaces mass manufacturing as the main source of wealth growth, and thereby reshapes the terms and conditions of conflict.

John Arquilla and David Ronfeldt identify the importance of networks in cyberspace and discuss a new type of societal warfare - netwar. Netwar refers to an emerging type of conflict at societal levels involving measures short of war, in which the protagonists use network forms of organisations, doctrine, strategy and communication (Arquilla & Ronfeldt 1996). In networks, actors generally consist of dispersed, often small groups who communicate, coordinate and act in a networked manner, without precise central leadership or headquarters. In contrast to traditional modes of conflict, decision-making may be deliberately decentralized and dispersed (Arquilla & Ronfeldt 2001).

Over the last years we have witnessed a rapid growth of social media platforms. People communicate, interact and socialize through social networks like Facebook, Twitter, Google+, Skype, Flickr, WordPress, YouTube and LinkedIn. Social media platforms have created a public sphere that provides information about every aspect of human life and enables the spread of ideas. Recent events in a number of countries like Iran, Tunisia and Egypt, have demonstrated the role that social media can play in mobilizing for collective action and political protest. Bloggers, web activists and average citizens, acting as online citizen-journalists, broadcasted amateur videos, audio interviews, tweets and emails, producing a near real-time flow of information that reached millions of users around the world (Liaropoulos 2013a).

The social wave paradigm focuses on the semantic layer of cyberspace and explores how the spread of information and thereby ideas, can trigger changes in identity and perceptions that challenge the existing understanding of war and conflict. Social wave scholars' interest on cyberspace is not on cyber-attacks per se, but rather on how (hostile) actions in cyberspace alter the way information societies perceive and conduct (cyber)war. Ideas do not threaten critical infrastructure assets, but can challenge the socio-political order by shaping the meaning and discourse of security. The social wave scholars focus on the use of social media networks in socio-political conflicts (Morozov 2011). The unique sociopolitical dynamics of the so-called Twitter Revolution and Facebook Revolution that took place in Iran, Tunisia and Egypt demonstrate that the effect of social media should not be taken for granted (Liaropoulos 2013a).

The control that is exercised on the Internet, the percentage and social characteristics of the population that uses social media, the ability of actors to disrupt the flow of information or to hack e-mail accounts and the role of the old media are only some of the myriad factors that have to be taken into consideration in order to value the role of ICTs in cyberconflicts. What are the conditions under which information sharing via social networks is successful in mobilizing the masses and translated into power?

3. Searching for middle ground: present gaps and future challenges

The above review vividly highlights the complexity of approaching the nature of conflict in the cyber domain. Part of the problem is rather conceptual. Defining concepts like war, security, power and conflict has always been puzzling. Adding the 'cyber' prefix has made this task even more challenging. Cyberspace is not a static domain, but a dynamic construct that is still rapidly expanding. In sharp contrast to land, sea, air and space, cyberspace is not part of nature; it is human-made and therefore shaped by human behaviour. Even though the cyber-terminology is a topic beyond the scope of this paper, it is self-explanatory that semantics are important. The way key concepts relating to cyberspace are understood, defines expectations and expectations are critical in shaping action. Cyberspace lacks 'a common language', meaning a set of concepts and terms that scholars and practitioners agree upon (Lieberthal & Singer 2012: 20). Another shortcoming in approaching cyberconflict is that up to a certain degree hard evidence is lacking. It is difficult to adequately assess the transforming nature of a new form of conflict, when the empirical evidence is lacking. Although cases like the cyber-attacks in Estonia and Georgia offer some insight on the nature of cyberconflicts, the evidence is still controversial and circumstantial.

Using the historical analogy of air power, it appears that in relation to the capabilities of cyber weapons we are in the same position, as air warfare was in the 1930s. Despite the developments in aviation technology and the use of air forces in World War I, there was no large scale deployment of air power before World War II. Therefore, any evaluation of the impact of air power in the 1930s would be ambiguous. Likewise, we are at the same point regarding cyberconflict. We have witnessed cyberspace becoming a domain of competition and confrontation for state and non-state actors, yet we have not experienced the result of a major use of cyber weapons against a society's critical infrastructure (Krepinevich 2012: ii). A closer look at the two paradigms reveals that both approaches, although from different angles, touch upon the central issue of power; or rather of cyber power. Both paradigms shed light on the understanding of cyberspace as an instrument of power. The discussion about the nature and typology of cyber power will only benefit the understanding of cyberspace. The debate is still in its infancy (Cavelty et.al 2007, Kramer et.al 2009, Betz & Stevens 2011, Klimburg 2011, Nye 2011, Langø 2013, Singh 2013) and due to the unique nature of cyberspace, has to be enriched from a multidisciplinary and interdisciplinary approach that combines theories from various academic areas (information technology, media and communication, international relations theories, security studies, conflict studies, social movement theories, sociology, etc). Many factors that are associated with cyber power, like the growth of the Internet, the level of participation in social network systems and the global governance of cyberspace are nonlinear. Cyberspace is diffuse and malleable and this will affect the way power is created, used and understood.

Future research on the area of cyber power should adopt a more systematic framework of analysis. How and to what extent does the use of cyberspace produce military, economic, political and social effects? What are the effects of power diffusion in cyberspace? How can actors utilize cyberspace in order to influence or compel other actors in order to achieve political goals; Since cyberspace is a new domain with unique characteristics, is it rational to conceive power in cyberspace in the same way as in the physical world? Anyone attempting to untangle the complex relation between cyberspace and power cannot afford to ignore these questions.

4. Conclusion

For more than two decades, experts from various fields have approached cyberspace and the security challenges that this new domain entails. The literature is vast and with noticeable shortcomings. This paper has attempted to frame the debate on cyberconflicts by identifying and categorizing the most influential works into two working paradigms. The (r)evolutionary paradigm and the social wave paradigm provide us with a broad understanding of the changes that ICTs brought about in the concepts of war, security and conflict. Despite the different approaches and the battle about the cyber-terminology, one cannot ignore the reality. Cyberspace is deeply embedded in our everyday life and is blurring the lines between public and private, national and international. Cyberspace is a distinct domain, with unique characteristics, where terms like war, security and conflict have still to be conceptualised. Whether such concepts that have been developed in the pre cyber era will be reconceptualised, remains to be seen. The review of the literature has demonstrated that the concept of cyber power is critical in addressing the potential impact of ICTs.

References

Arquilla, J. & Ronfeldt, D. (1996) *The Advent of Netwar,* RAND, National Defence Research Institute, Santa Monica CA.

Arquilla, J. & Ronfeldt, D. eds (1997) *In Athena's Camp: Preparing for Conflict in the Information Age*, RAND, National Defense Research Institute, Santa Monica, CA.

Arquilla, J. & Ronfeldt, D. eds (2001) *Networks and Netwars: The Future of Crime, Terror and Militancy*, RAND, National Defense Research Institute Santa Monica CA.

Barzashka, I. (2013) 'Are Cyber-Weapons Effective?' *The RUSI Journal,* Vol 158, No.2, pp.48-56.

Betz, D. & Stevens, T. (2011) *Cyberspace and the State. Toward a Strategy for Cyber-Power,* Adelphi Series 424, Routlegde, IISS, Oxon.

Blank, S. (2008) 'Web War I: Is Europe's First Information War a New Kind of War?', *Comparative Strategy,* Vol 27, No.3, pp.227-247.

Brito, J. & Watkins, T. (2011) 'Loving the Cyber Bomb? The Dangers of Threat Inflation in Cybersecurity Policy', *Harvard National Security Journal,* Vol 3, No.1, pp.39-84.

Bunker, R. (1996) 'Advanced Battlespace and Cybermanoeuvre Concepts: Implications for Force XXI', *Parameters,* Vol 26, No.3, pp.108-120.

Cavelty, M. (2007) *Cyber-Security and Threat Politics: U.S Efforts to Secure the Information Age,* Routledge, New York.

Cavelty, M. et.al (2007) *Power and Security in the Information Age: Investigating the role of the state in cyberspace,* Ashgate, Burlington.

Choucri, N. (2012), *Cyberpolitics in International Relations,* The MIT Press, Cambridge Massachusetts.

Clark, W. & Levin, P. (2009) 'Securing the Information Highway', *Foreign Affairs*, Vol 88, No.6, pp.2-9.
Clarke, R. & Knake, R. (2010) *Cyber War: The next threat to national security and what to do about it,* Harper Collins Publishers, New York.
Costigan, S. & Perry, J. eds (2012) *Cyberspaces and Global Affairs,* Ashgate, Surrey.
Farwell, J. & Rohozinski, R. (2011) 'Stuxnet and the Future of Cyber War', *Survival*, Vol 53, No.1, pp.23-40.
Gartzke, E. (2013) 'The Myth of Cyberwar. Bringing War in Cyberspace Back Down to Earth', *International Security,* Vol 38, No.2, pp.41-73.
Gray, C.S. (2001) Weapons *for Strategic Effect: How Important is Technology?* Occasional Paper, No.21, Center for Strategy and Technology, Air War College, Maxwell AFB, AL.
Gray, C.S. (2013) *Making Strategic Sense of Cyber Power: Why the sky is not falling,* Strategic Studies Institute and U.S Army War College Press, Carlisle, PA.
Hussain, M. & Howard, P. (2013) 'What best explains successful protest cascades? ICT and the fuzzy causes of the Arab Spring', *International Studies Review*, Vol 15, No.1, pp.48-66.
Karatzogianni, A. (2006) *Politics of Cyber Conflict,* Routlege, London and New York.
Karatzogianni, A. ed. (2009) *Cyber Conflict and Global Politics*, Routledge, London and New York.
Kerry, J. (2013) *Nomination: U.S. Senate Committee on Foreign Relations,* Washington D.C.
Klimburg, A. (2011) 'Mobilizing Cyber Power', *Survival,* Vol 53, No.1, pp.41-60.
Knox, M. & Murray, W. eds (2001) *The Dynamics of Military Revolution, 1300-2050,* Cambridge University Press, Cambridge.
Kramer, F. et.al (2009) *Cyberpower and National Security,* Potomac Books & National Defense University Press, Washigton D.C.
Krepinevich, A. (2012) *Cyber Warfare: A Nuclear Option?,* Washington D.C, Center for Strategic and Budgetary Assessments.
Langø, H.I. (2013) 'The Limits of Compulsory Cyber Power: Assessing Ecological Potential and Restraints in the Digital Domain', *NUPI Working Paper 819*, Oslo, Norwegian Institute of International Affairs.
Langø, H.I. (2013a) 'Slaying Cyber Dragons: Competing Academic Approaches to Cyber Security', *NUPI Working Paper 820*, Oslo, Norwegian Institute of International Affairs.
Lawson, S. (2011) 'Beyond Cyber-Doom: Cyberattack Scenarios and the Evidence of History', *Working Paper,* No.10-77, Mercatus Center, Fairfax, VA.
Liaropoulos, A. (2011) 'Power and Security in Cyberspace: Implications for the Westphalian state system', in *Panorama of Global Security Environment*, Centre for European and North American Affairs, Bratislava.
Liaropoulos, A. (2013) 'Exercising state sovereignty in cyberspace: An international cyber-order under construction?', *Journal of Information Warfare,* Vol 12, No.2, pp.19-26.
Liaropoulos, A. (2013a) 'The Challenges of Social Media Intelligence for the Intelligence Community', *Journal of Mediterranean and Balkan Intelligence,* Vol 1, No.1, pp.5-14.
Liaropoulos, A. (2014) 'Cyberspace, Sovereignty and International Order, *International Relations and Security Network,* 30 January, [online], http://www.isn.ethz.ch/Digital-Library/Articles/Detail/?lng=en&id=176144.
Libicki, M. (1995) *The Mesh and the Net*, Institute for National Strategic Studies, National Defence University, Washington D.C.
Libicki, M. (2007) *Conquest in Cyberspace. National Security and Information Warfare*, Cambridge University Press, Cambridge.
Lieberthal, K. & Singer, P. (2012) *Cybersecurity and U.S.-China Relations,* Brookings, Washington D.C.
Lonsdale, D. (2004) *The Nature of War in the Information Age: Clausewitzian Future*, Frank Cass, London.
Lynn, W.J III (2010) 'Defending a New Domain', *Foreign Affairs*, Vol 89, No.5, pp.97-108.
Morozov, E. (2011) *The Net Delusion. The Dark Side of Internet Freedom*, Public Affairs, New York.
Nye, Joseph (2011) *The Future of Power,* Public Affairs, New York.
Rattray, G. (2001) *Strategic Warfare in Cyberspace*, The MIT Press, Cambridge MA.
Rid, T. (2012) 'Cyber War will not take place', *Journal of Strategic Studies*, Vol 35, No.1, pp.5-32.
Rovner, J. (2013) 'Intelligence in the Twitter Age', *International Journal of Intelligence and Counterintelligence,* Vol 26, No.2, pp.260-271.
Schwatz, J. (2007) 'Preparing for a Digital Pearl Harbor', *The New York Times,* 24 June, [online], http://www.nytimes.com/2007/06/24/business/worldbusiness/24iht-cyber.1.6299676.html?_r=0.
Singh, J.P. (2013), 'Information Technologies, Meta-Power and Transformations in Global Politics', *International Studies Review,* Vol 15, No.1, pp.5-29.
Tabansky, L. (2011) 'Basic Concepts in Cyber Warfare' *Military and Strategic Affairs*, Vol 3, No.1, pp.75-92.
Toffler, A. & Toffler H. (1993) *War and Anti-War: Survival at Dawn of the 21st Century,* Warner Books, New York.
Toffler, A. (1980) *The Third Wave,* Bantam, New York.
Tsarougias, N. (2012) 'Cyber attacks, self-defense and the problem of attribution', *Journal of Conflict & Security Law,* Vol 17, No.2, pp.229-244.
Van Creveld, M. (1991) *Technology and War. From 2000 B.C. to the Present*, The Free Press, New York.
Williams, R.D. (2001) 'Is the West's Reliance on Technology the Panacea for Future Conflict or its Achilles heel?', *Defence Studies*, Vol 1, No.2, pp.38-56.

Planning Method of Information Security for Military Organizations[1]

José Martins[1], Henrique dos Santos[2], Mendes Dias[1] and José Borges[1]
[1]Academia Militar – CINAMIL, Lisboa, Portugal
[2]Universidade do Minho – DSI, Guimarães, Portugal
jose.carloslm@gmail.com
hsantos@dsi.uminho.pt
mendes.dias@academiamilitar.pt
jose.borges@academiamilitar.pt

Abstract: The main question to answer is: how to ensure the confidentiality, integrity, and availability of information within a military organization in information warfare environment, in order to minimize the risk of information security? We propose a planning method, guided by some principles of war and taking into account known modes of action of an enemy. The planning method enables the identification of: (i) the main methods of attack that can occur; (ii) the controls' baseline applied in military organizations; (iii) the security controls applied by attack methods and validation the their effectiveness, according to the specific method of attack. In order to put into practice the designed planning method of information security, the authors followed an interpretive, qualitative, and inductive epistemological orientation, which is used as the main research method for the Analysis of Content, the Focus Group and the Case Study. The proposed method of planning is based on: (i) a model that allows the identification of possible methods of attack to information, carried out using the vectors of Physical, Human and Technological infrastructure attack; (ii) a framework of categories of information security controls (security dimensions: Organizational, Physical and Environmental, Human, and Technological); (iii) a matrix of decision support sustained by the attack vectors of an opponent, and the possible effects of information security controls (Prevent, Detect, Deter, Deflect, Recover or React). The method purposed in this work identifies the best combination of security controls to be applied against a particular method of attack, taking into account the lessons learnt. It is also based on the selection and recovery of the solutions successfully applied to past cases. This allows the military decision-maker to plan, focusing on the modes of action of an opponent, while considering the principles of war adapted to information security (Economy of Force, Maneuver, Unit of Command and Offensive).

Keywords: information security planning, information security method, methods of attack, information security framework

1. Introduction

In organizations in general and in the military in particular, information is one of the most important assets of support to all business processes. Its manipulation depends on three main elements: (i) technology, which allows the storage, the processing, and transmission; (ii) stakeholders, who can access it through the Internet or private networks and; (iii) the business process that use it (Posthumus & Von Solms, 2004). Thus, it´s essential that we seek, continuously, to ensure the fundamental properties of security, such as confidentiality, integrity, and availability (ISO/IEC27001, 2005; Posthumus & Von Solms, 2004; Siponen & Oinas-Kukkonen, 2007).

The effects of some of the methods of attack supported primarily on Information Technology (IT), can be depicted in the cyber attack launched against Estonia in April and May 2007, which led to the shutdown of several State activities (Tikk, 2008) or in the conflict in Georgia in 2008 (Tikk, et al., 2008). Moreover, the demonstrated ability of some countries to conduct *Cyber Warfare and Computer Network Operations* (Andress & Winterfeld, 2011; Carr, 2012; Krektel, Bakos, & Barnett, 2009; TRADOC-PAM525-7-8, 2010), completely justified new approach to information security in military organizations.

Computer attacks, such as Malware, the Denial of Service, the Packet Sniffer, the Masquerade, and the Man-in-the-Middle have been identified as main methods of attack focused on computer networks (Kurose & Ross, 2010). The majority of these methods are employed through the Internet (Libicki, 2007; Stolfo, et al., 2008; Tipton & Krause, 2007) and allow exploiting vulnerabilities, resulting from software (Correia & Sousa, 2010; Goertzel, et al., 2007; Pfleeger & Pfleeger, 2007). Other possible methods of attack, which should also be a concern for the military decision-makers are those supported in acts of Social Engineering (Hadnagy, 2011; Hogben, 2007; Mann, 2008), aimed at the manipulation of the human element. Other classic methods are

[1] This study does not reflect the vision or institutional guidelines of the Portuguese Army doctrine.

based in kinetic actions that can lead to the physical destruction of facilities, equipment or critical infrastructure that support the activities of organizations (Erbschloe, 2005).

The increasing importance of Information Security in military organizations is also due to the emergence of concepts, developed especially in the last two decades, were referenced predominantly, in the military, such as *Information Superiority* (Alberts, Garstka, Hayes, & Signori, 2001) and *Information Warfare* (Arquilla & Ronfeldt, 1999; M. Libicki, 1995; Waltz, 1998), based mainly, on the idea that information is seen as simultaneously a weapon and a target (Hutchinson, 2003).

Due to the concepts of military doctrine, the use of IT as an offensive weapon, the emergence of cyberspace as a new dimension of the battlefield, and the importance of information in achieving information superiority in an Information Warfare (IW) environment, it is important to develop new approaches to information security and hence new processes for planning in military organizations. This activity looking for new approaches can be seen as a *"Defensive Battle"* (Chesla, 2004), framed in IW. The context described justifies the relevance of *designing* new information security planning method, for military organizations, assuming as essential the focus on the possible modes of action of an opponent (i.e. their methods of attack).

The main question to answer is: how to ensure the confidentiality, integrity, and availability of information within a military organization in the context of an information warfare environment, in order to minimize the risk of information security?

This paper, in order to answer the question, is divided into four sections. The first section presents the problem and sets out the main objective of the study. The second gives us, briefly, the applied research plan. The third section points out the main results of the study and finally to conclude, the fourth section will present the findings and the limitations of the study.

2. Research plan

In order to put into practice the designed planning method of information security, we followed an interpretive, qualitative, and inductive epistemological orientation, which is used as the main research method for the Analysis of Content (Bardin, 2011), the Focus Group (Liamputtong, 2011) and the Case Study (Remenyi, 2012).

The *Content Analysis* takes into account the main international standard of information security management (ISO/IEC27001, 2005), as well as the military security standards that have been applied in the Portuguese Army, the security doctrine of NATO, and the Certified Information System Security Professional (Harris, 2008). The *Focus Group* consisted of a panel of Portuguese military specialists and civil personnel from different areas of expertise in information security (i.e. computer networks, physical security). A *case study* was conducted in a military organization with level one as classification; an organization that has vital information for the accomplishment of the mission of the Portuguese Army, which is necessary to ensure the security (Martins, Santos, Rosinha, & Valente, 2013). In addition to the methods of research set, we used, predominantly scientific knowledge related with the academic discipline of *Security of Computer Networks*, *Software Engineering*, and *Software Security*.

The absence of a general theory of information security required the development of a matrix of concepts presented in Table 1.

Table 1: Matrix of concepts

Concepts	Definition
Information	Existing data in any form and shape, which is assigned a meaning, after having been organized in a useful manner, in order to impart a meaningful message to the recipients of its use (JP3–13, 2006; Laudon & Laudon, 2006).
Attacker	Individual or group of individuals, internal or external to the organization that with the execution of one or more methods of attack, seek to aim the fundamental properties of information security in to order to accomplish an operational objective (FM100-06, 1996; Howard & Longstaff, 1998; Pfleeger & Pfleeger, 2007).
Threat	Potential cause of an incident of information security, which can result in damage to the system or organization (ISO/IEC13335-1, 2004).

Concepts	Definition
Attack Method	Action or set of actions, supported in certain tools, which are used to exploit the vulnerabilities of the main components of the IS, with the purpose of, directly or indirectly, achieving the security properties of information and consequently producing operational effects (Howard & Longstaff, 1998).
Action	Activity that causes an event in a system (i.e. application) and the possible change of its state (Howard & Longstaff, 1998). In terms of information security, actions may cause information security incidents.
Tool	Means, weapons or resources that are used to exploit the vulnerabilities of critical assets of an organization, i.e. the targets (Howard & Longstaff, 1998).
Target	Logical entities (i.e. account, information/data); physical entities (i.e. computer, network); human resources of the organization (i.e. decision-makers, experts); means of transmitting information (i.e. the wiring, electromagnetic radiation, sound waves); physical infrastructure (i.e. facilities, data centre, meeting rooms), or rather, all of the critical assets of an organization which achieve directly or indirectly, confidentiality, integrity and availability of information (AAP-6, 2009; Howard & Longstaff, 1998).
Vulnerabilities	Characteristics of critical assets (targets) of an organization, which consist of weaknesses that can be exploited by an attacker to execute a method of attack (ISO/IEC27001, 2005).
Information Security Property	Inherent quality of information, which indicates to what extent that property of security, exists in that information (ISO/IEC27001, 2005).
Operational Effects	Main purposes or objectives of military nature to achieve by an opponent, in order to contribute to information superiority in an IW environment.
Security Controls	Resources to be used to reduce or remove the vulnerabilities of an organization's critical assets, that may be done by policies, processes, procedures, best practices or security mechanisms. Can be applied to prevent, detect, deter, deflect, recover or react to one or more methods of attack (Dhillon, 2007; Pfleeger & Pfleeger, 2007).

3. Research outcomes

The proposed method of planning is based on: (i) a model that allows for the identification of possible methods of attack to information, carried out using the vectors of Physical, Human and Technological infrastructure attack (Martins, Santos, Nunes, & Silva, 2012b); (ii) a framework of categories of information security controls (security dimensions: Organizational, Physical and Environmental, Human, and Technological) (Martins, Santos, Nunes, & Silva, 2012a); (iii) a matrix of decision support sustained by the attack vectors of an opponent, and the possible effects of information security controls (Prevent, Detect, Deter, Deflect, Recover or React); (iv) Organizations analysis process (Figure 1).

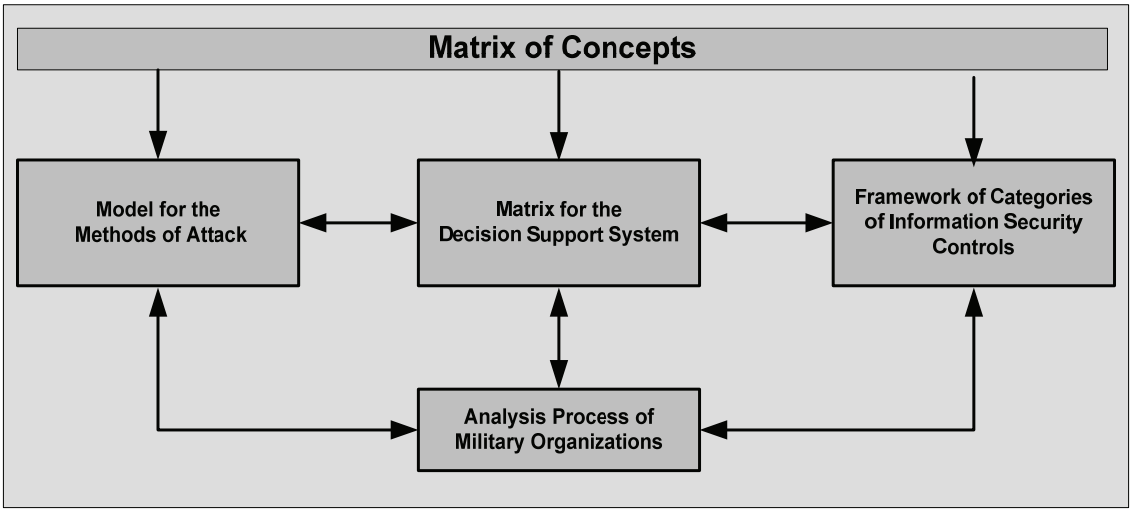
Matrix of Concepts
Model for the Methods of Attack
Matrix for the Decision Support System
Framework of Categories of Information Security Controls
Analysis Process of Military Organizations

Figure 1: Main components of the planning method of information security

The main objective of the planning method is to assist the military decision makers to select the best combination of security controls to be implemented against a practical method of attack to information. The design of information security planning method is built conforms the context model shown in Figure 2.

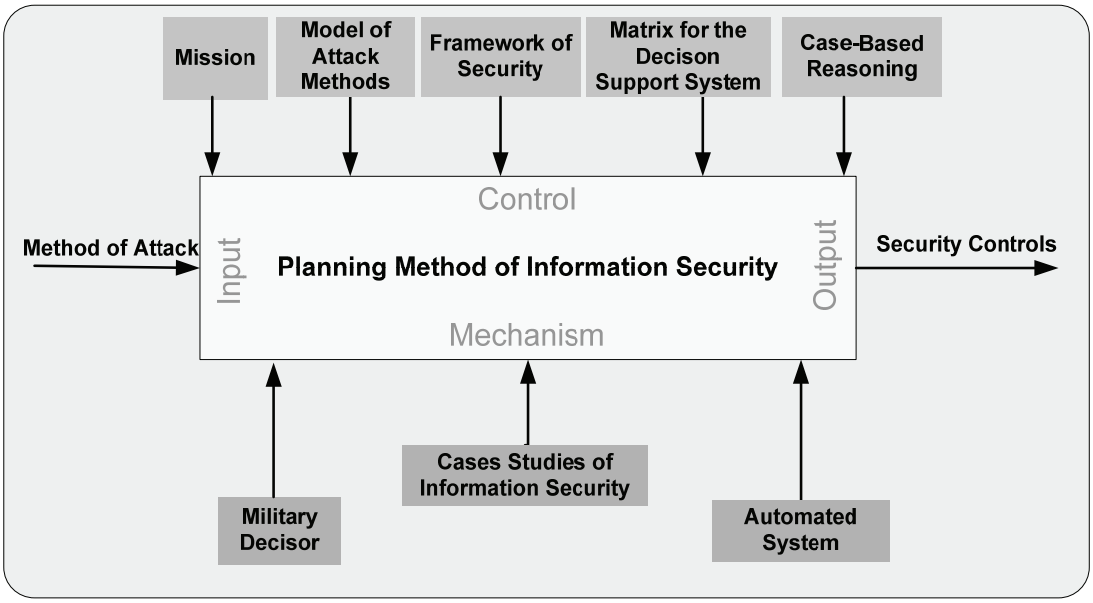

Figure 2: Context model for the information security planning method

Faced with a likely method of attack held by the enemy and taking into account the mission of friendly forces, the method of planning proposed allows, through the integration of the model of attack methods and the *framework* of information security categories, to support the decision of selecting the best set of security controls, to be implemented, without forget the experience of experts.

According to the Portuguese Army military doctrine, planning "is the process by which a commander visualizes an end state, conveys an effective method of reach and communicates his vision, intent and decisions" (PDE5-00, 2007, p. 21).

The military commander must determine / guess the possible actions of the enemy, in order to organize friendly forces, considering permanently the principles of war (PDE2.09.00, 2010; RC130-1, 2005). In planning, the military decision maker identifies the possible hazards to the accomplishment of the mission, assessing their severity and probability of occurrence and communicates through policy, measures to be implemented to minimize them (PDE5-00, 2007). The result of planning is a plan that facilitates action and which is constantly revised.

To execute the *design* of the planning method, the following four propositions are considered:

- The information security is built on the basis of possible methods of attack of an opponent, carried out in the attack vector: physical, human and technological infrastructure (Martins, et al., 2012b).
- The security controls are integrated into the major categories of information security, according to the dimensions of security: Organizational, Physical and Environmental, Human and Technological (Martins, et al., 2012a).
- The desired effects with the security controls of information implemented are to prevent, detect, deter, deflect, recover or react to an attack method (Dhillon, 2007; Pfleeger & Pfleeger, 2007).
- The planning of information security is based on lessons learned (PDE0.32.00, 2012), or rather aims at selecting, retrieving and adapting ongoing solutions attack methods or information security incidents that have occurred, that resemble the situation.

The *design* of information security planning method takes into account, given the specificity of the military organization, the most important principles of war applied in the planning and conduct of military operations (Couto, 1988; RC80-5, 1991; RC130-1, 2005). The principles of war are under military laws, which were for centuries, and are still considered in military doctrine, as the key elements in operational planning, being fundamental to the characterization of the specificity of the planning of military operations (Couto, 1988). The adaptation of these principles to the planning of information security method is illustrated in Table 2.

Table 2: Principle of war adapted in information security

Principles of War	Goal
Economy of Force	The information security controls implemented must protect the greatest number of attack methods and ensure maximum possible effects (i.e. prevent, recover).
Maneuver	The controls implemented should ensure protection of the main vector of attack, defense in depth and mutual support between them.
Unity of Command	The method of planning should ensure integration of all management levels in military organizations, through a common sight in the planning, implementation (includes maintenance, monitoring and audits) and cohesion of controls applied.
Security	The planning method should allow planning actions to obtain information about an opponent in order to anticipate its modes of action.
Offensive	The planning method should allow offensive action within the information security, rather than react. Simultaneously should allow anticipating the possible methods of attack an opponent and exploit its vulnerabilities.

The planning method enables supporting and identifying the best combination of security controls to be implemented, to deal with a certain method of attack. As depicted in Figure 3, the instantiation of the possible methods of attack to a specific organization is performed on the basis of a model of attack methods (Martins, et al., 2012b).

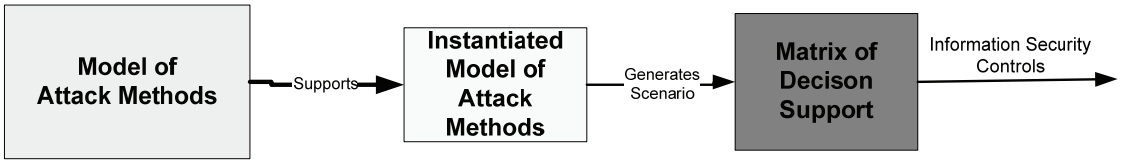

Figure 3: Scenario and decision support matrix to information security

The planning of the selection of controls to implement will be based on the decision support matrix shown in Figure 4. This has two axes, the dimensions of information security (according to the attack vectors) and the effects of information security controls.

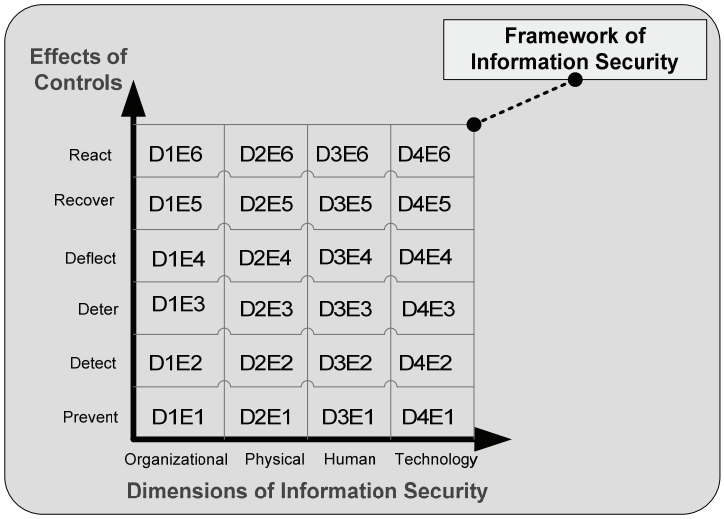

Figure 4: Matrix for the decision support system in information security

Moreover, in the identification and selection of information security controls to be applied, a *framework* of categories of information security controls (Martins, et al., 2012a) and decision support matrix are utilized (Figure 4). Information security controls were found in the major categories, whose efficiency, in comparison

to a particular attack method is validated, based on the lessons learned in the military organization or in other organizations. Thus, the need for collaboration and information sharing between military organizations is justified.

It is important to consider in the identification and selection of these security controls the following criteria for its implementation: (i) control is necessary to ensure the protection of one or more dimensions of information security against possible attack methods used by an opponent. This is because control of information security can be implemented to protect information from one or more methods of attack and possibly with different goals; (ii) control is specific (single) and should be measured qualitatively or quantitatively; (iii) the implementation of control is achievable within a time period acceptable to the organization, and realistic when compared to the criteria defined by the military organization.

The controls that are identified as fundamental for information security, due to the possible occurrence of the most likely and dangerous attack methods to the organization, and should be applied in military organizations according to the specificity of its management which, in most cases, is done according to rules, processes or standards for permanent implementation. This interaction between a decision support system for information security and military organizations (Figure 5) is possible at a later stage, after the application of the selected controls and the collection of information security lessons learned. These will allow validating the effectiveness of information security controls applied across the organization.

To finish drawing the planning method of information security, it is necessary to sustain the implementation of controls on lessons learned. This aim is achieved through the selection and recovery of solutions implemented successfully (i.e., combinations of information security controls) to past cases of attack methods that resemble the situation under study, adapting them to the new methods of attack. Therefore, the process of analyzing the military organization referenced in Figure 5 [2] yields the *Case Studies* that support the planning of information security process and simultaneously, provide feedback to the model of attack methods and framework of categories of information security controls.

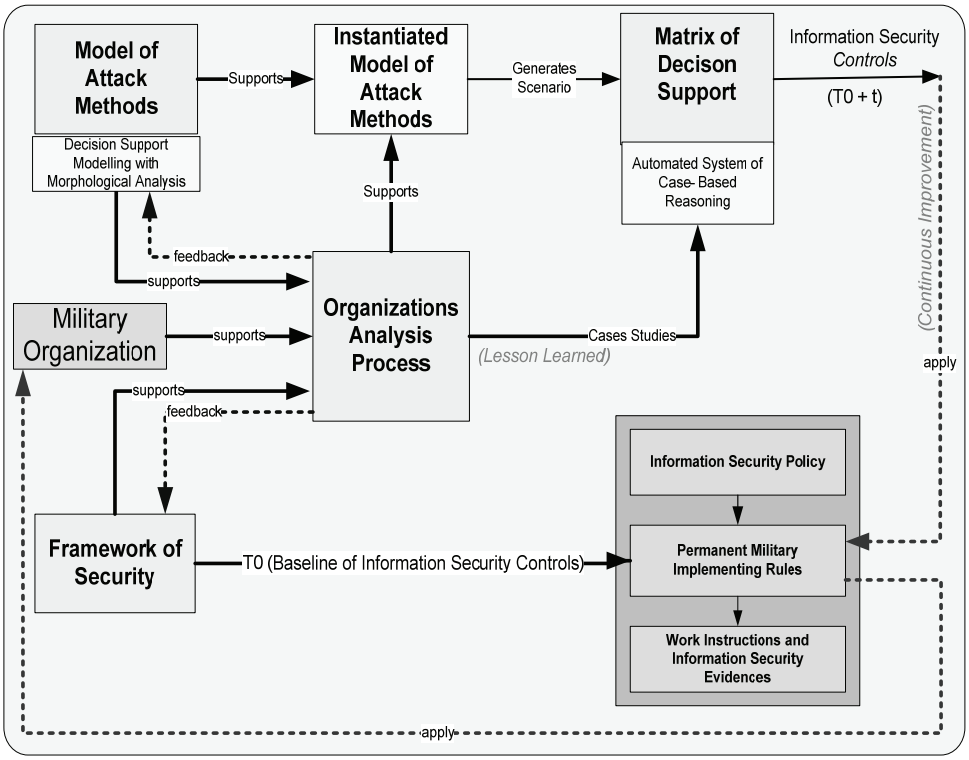

Figure 5: Method of information security planning

[2] At T0 (at the beginning of the use of the method), there is already a set of controls applied in military organization, which should form the baseline security organization for the main attack methods that will change over time (T0 + t).

The method of planning information security is implemented in a military organization according to the process illustrated in Table 3, which description matches the requirements of the quality management systems oriented by ISO/IEC 9001 (Martins, et al., 2013).

Table 3: Overview of the information security planning process

Goal	Selection of the combinations of security controls of information more efficient, to deal with an enemy attack, and definition of the baseline information security to be implemented in order to avoid in the IW environment the information superiority of the adversary
Main Inputs	Model of attack methods Framework of categories of information security controls Matrix of decision support system in information security Process analysis of a military organization
Main Outputs	Baseline security controls for information Information security controls by method of attack
Process Control	Documental evidence of the operational procedures Reports of Case Studies with lessons learned in military organizations where the method is applied Reports of planned and unannounced audits
Main Documentation	Methods of attack and information security framework (reality) Plan information security Policy information security Technical information security policies Operational procedures Registration of the evidence of the case and the controls implemented
Coordination	Performed by the chain of command of the military organization
Necessary Resources	Automated System of Case-Based Reasoning (“database of past cases”) Experts in the fields of information security Automated process of gathering evidence of information security in military organizations (i.e., incidents)
Responsible for the Process	Security officer by authority delegation of the Commander in the military organization

The method of information security planning for military organization in an IW environment:

- Considers the principles of war and military security (“need to know”, “least privilege”, “responsibility, loyalty and trust” and “the defense in depth”).
- Facilitates its operationalisation through the chain of command, by: (i) centralized planning and efforts orientation by attack method; (ii) all organization levels; (iii) security capabilities (dimensions and categories of controls) and skills of employees (security controls).
- Allows an agile management that enables simulate and anticipate attack methods (interconnection with the model of attack models), and facilitate the inclusion of new categories and information security controls.
- Supports the management of information security lessons learned concerning to incidents of information security, taking into consideration that there is no single “recipe” of information security for all military organizations.
- And finally, enables all employees to raise awareness for information security, through a single view, i.e. shared model for all employees.

This planning method respect the principle of *Unity of Command* by ensuring the integration of management levels of the military organization, which provides a common view of information security to all employees and “cohesion” of measures to be applied. The proposed planning method is also guided by the principles of *Security and Offensive*; the model of attack methods allows, both friend and foe, planning actions of obtaining information and offensive.

The method of information security planning supported by the propositions already focused and the principles of war, allows answering the following operational issues: (i) How can a method of attack on information from a military organization to contribute to an opponent achieve information superiority? (ii) What are the targets that a particular method of attack can explore and a determined target that can be exploited by attack

methods? (iii) What is the minimum *baseline* of information security controls to military organizations? (iv) What is the contribution of a particular security check to protect the military organization of a particular attack method or set of methods of attack to information?

4. Conclusions

The method proposed in this work contributes for the identification of the best combination of security controls to be applied against a particular method of attack, is used taking into account, also the lessons learnt. It is based on the selection and recovery solutions successfully applied to past cases. This allows for the military decision-maker to plan, focusing on the modes of action of an opponent, while considering the principles of war adapted to information security (Economy of Force, Maneuver, Unit of Command, Security and Offensive).

The main outcomes of this work can be summarized as follows: (i) a model of attack methods to information; (ii) a *framework* of categories of information security controls; (iii) a matrix for the decision support system for planning the selection of information security controls to implement within the military organization; (iv) a method of information security planning, for military organizations to perform the management of the lessons learned, enabling continuous improvement.

However, the integration of multiple subject areas and the need for validation of the components of the method, points out some limitation in our study, which are listed below.

The model of attack methods proposed to identify all possible attack methods needs to implement the method of *Morphological Analysis General* (Ritchey, 2011) of the establishment of the inconsistency matrix, which validates all possible intersections of the conditions of the model variables.

There is also the need to define and validate rigorously the descriptors of the variables and conditions of the methods of attack model, as well as the categories of information security *framework*.

In this study no controls of the categories of information security *framework* were identified, nor associated metrics of efficiency, and work guidelines to the possible controls, as they fall outside the scope of this work.

The study relies on a single Case Study based on a Organization of the Portuguese Military Army, which is certainly insufficient; since it leads to that eventually some particular aspects have been forgotten. We tried to reduce this possibility with the analysis of all documentation of military security of the Portuguese Army and depth given to the Case Study.

The gathering and processing of empirical data in the context of information security in military organizations of the Portuguese Army, to operationalize the planning method, justifies the need for an automated system able to encode and store the known Case Studies. This “database of knowledge” will allow the military decision maker to identify the best combination of information security controls to be implemented for the all possible attack methods to information.

There is one aspect not covered in this study, which is critical for the future application of the method of planning, which is the development of an automated decision support system based on *Case Based Reasoning* (Kolodner, 1992).

Information security in military organizations in an IW environment is intended to ensure the fundamental properties of information security, in order to contribute to information superiority. The selection of controls to implement should be taken: (i) according to the methods of the enemy’s attack; (ii) by the integration of the dimensions of security; (iii) and considering the lessons learned.

Acknowledgements

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: PEst-OE/EEI/UI0319/2014.

Reference

AAP-6. (2009). NATO Glossary of Terms and Definitons.
Alberts, D., Garstka, J., Hayes, R., & Signori, D. (2001). Understanding Information Age Warfare, CCRP Publication Series, Washington, United States of America.
Andress, J., & Winterfeld, S. (2011). Cyber Warfare: Techniques, Tactics and Tools for Security Practitioners: Syngress Media Inc.
Arquilla, J., & Ronfeldt, D. (1999). The advent of netwar: Analytic background. Studies in Conflict & Terrorism, 22(3), 193-206.
Bardin, L. (2011). Análise de Conteúdo: Edições 70, LDA. (in Portuguese)
Carr, J. (2012). Inside Cyber Warfare (Second Edition ed.): O´Reilly.
Chesla, A. (2004). Information Security: A Defensive Battle. Information Security Journal: A Global Perspective, 12(6), 24-32.
Correia, M. P., & Sousa, P. J. (2010). Segurança no Software. Lisboa: FCA. (in Portuguese)
Couto, A. C. (1988). Elementos de Estratégia (Vol. I). Lisboa: Instituto de Altos Estudos Militares. (in Portuguese)
Dhillon, G. (2007). Principles of Information Systems Security - Text and Cases: WILEY.
Erbschloe, M. (2005). Physical Security for IT. Elsevier Digital Press, United States of America.
FM100-06. (1996). Information Operations. Washington: Headquarters, Department of the Army.
Goertzel, Winograd, McKinley, Oh, Colon, McGibbon, et al. (2007). Software Security Assurance: A State-of-the-Art Report: Defense Technical Information Center.
Hadnagy, C. (2011). Social Engineering - The Art of Human Hacking: Wiley Published
Harris, S. (2008). CISSP All-in-One Exam Guide, Fourth Edition, McGraw-Hill, New York, United States of America.
Hogben, G. (2007). Security Issues and Recommendations for Online Social Networks. ENISA Position Paper No. 1: October.
Howard, J. D., & Longstaff, T. A. (1998). A common language for computer security incidents. Sandia Report: SAND98-8667, Sandia National Laboratories, http://www. cert. org/research/taxonomy_988667. pdf.,
Hutchinson, W. (2003). The Changing Nature of Information Security. Paper presented at the 1st Information Security Management Australian.
ISO/IEC13335-1. (2004). Information technology- Security techniques-Management of information and communications technology security. Part 1: Concepts and models for information and communication technology security management.
ISO/IEC27001. (2005). Information technology – Security techniques – Information Security Management Systems - Requirements.
JP3–13. (2006). Joint Doctrine for Information Operation, United States of America.
Kolodner, J. L. (1992). An introduction to case-based reasoning. Artificial Intelligence Review, 6(1), 3-34.
Krektel, B., Bakos, G., & Barnett, C. (2009). Capability of the People's Republic of China to Conduct Cyber Warfare and Computer Network Exploitation, Northrop Grumman Corporation, Report, United States of America.
Kurose, J. F., & Ross, K. W. (2010). Computer Networking, Addison Wesley, 5th Edition, United States of America.
Laudon, K. C., & Laudon, J. P. (2006). Management Information Systems, Prentice Halll, 9ª ed, United States of America.
Liamputtong, P. (2011). Focus Group Methodology - Principles and Practice: SAGE.
Libicki. (2007). Conquest in Cyberspace: National Security and Information Warfare: Cambridge University Press.
Libicki, M. (1995). What is Information Warfare? Washington: National Defense University.
Mann, I. (2008). Hacking the human: social engineering techniques and security countermeasures: Gower Publishing Company.
Martins, J., Santos, H., Nunes, P., & Silva, R. (2012a). Framework de Gestão de Segurança da Informação para Organizações Militares Orientada pelos Principais Vetores de Ataque. Paper presented at the Associação Portuguesa de Sistemas de Informação, Universidade do Minho, Guimarães, Portugal. (in Portuguese)
Martins, J., Santos, H., Nunes, P., & Silva, R. (2012b). Information Security Model to Military Organizations in Environment of Information Warfare. Paper presented at the 11 th European Conference on Information Warfare and Security, Laval, France.
Martins, J., Santos, H., Rosinha, A., & Valente, A. (2013). Information Security Military Standards versus ISO 27001 - A Case Study in a Portuguese Military Organization. Paper presented at the European Conference on Information Warfare and Security, Jyväskylä, Finlândia.
PDE0.32.00. (2012). Lições Aprendidas: Ministério da Defesa Nacional/Exército Português. (in Portuguese)
PDE2.09.00. (2010). Estudo do Espaço de Batalha pelas Informação (IPB): Ministério da Defesa Nacional / Exército Português. (in Portuguese)
PDE5-00. (2007). Planeamento Táctico e Tomada de Decisão: Ministério da Defesa Nacional - Exército Português - Comando da Instrução e Doutrina. (in Portuguese)
Pfleeger, C. P., & Pfleeger, S. L. (2007). Securiy in Computing, Prentice Hall, 4ª ed, United States of America.
Posthumus, S., & Von Solms, R. (2004). A framework for the governance of information security. Computers & Security, 23(8), 638-646.
RC80-5. (1991). Brigada de Infantaria Independente. Lisboa: Estado - Maior do Exército. (in Portuguese)
RC130-1. (2005). Regulamento de Campanha - Operações. Lisboa: Estado - Maior do Exército. (in Portuguese)
Remenyi, D. (2012). Case Study Research: Academic Publishing International.

Ritchey, T. (2011). Wicked Problems - Social Messes - Decision Support Modelling With Morphological Analysis (Vol. 17): Springer Verlag.

Siponen, & Oinas-Kukkonen. (2007). A review of information security issues and respective research contributions. ACM SIGMIS Database, 38(1), 60-80.

Stolfo, Bellovin, Hershkop, Keromytis, Sinclair, & Smith. (2008). Insider Attack and Cyber Security: Beyond the Hacker: Springer-Verlag New York Inc.

Tikk, E. (2008). National Defense Policies for Cyber Space – Background and Effect of the Estonian Cyber Attacks (Cooperative Cyber Defence Centre of Excellence). Academia Militar, Lisboa, Portugal.: .

Tikk, E., Kaska, K., Rünnimeri, K., Kert, M., Talihärm, A.-M., & Vihul, L. (2008). Cyber Attacks Against Georgia: Legal Lessons Identified, NATO Unclassified Report v1.0. Tallin, Estonia: Cooperative Cyber Defense Centre of Excellence.

Tipton, H., & Krause, M. (2007). Information Security Management Handbook (Sixth Edition ed.): Auerbach Publication.

TRADOC-PAM525-7-8. (2010). Cyberspace Operations - Concept Capability Plan 2016-2028: The United States Army's.

Waltz, E. (1998). Information Warfare: Principles and Operations, Artech House.

Comparison of two Specifications to Fulfill Security Control Objectives

Riku Nykänen andTommi Kärkkäinen
University of Jyväskylä, Jyväskylä, Finland
riku.t.nykanen@student.jyu.fi
tommi.karkkainen@jyu.fi

Abstract: Assuring information security is a necessity in modern organizations. Many recommendations for information security management (ISM) exist, which can be used to define baseline of information security requirements ensuring that an organization has implemented the selected practices. ISO/IEC 27001 prescribes a process for ISM system and guidance to implement security controls is provided in ISO/IEC 27002. Finnish National Security Auditing Criteria (KATAKRI) has been developed by the national authorities in Finland as tool to verify maturity of information security practices and hence it is organized in a form of compliance questionnaire. KATAKRI defines both security control objectives and absolute security controls to meet an objective. ISO/IEC 27001 requires selection of valid security controls whereas KATAKRI may force organization to implement controls that are not feasible from risk management or cost-benefit ratio point of view. In our work, we study the security control objectives and the actual controls of ISO/IEC 27002 and KATAKRI to analyze similarities and differences between the two specifications. Results can be used to determine in which sense KATAKRI could be used as ISO(IEC 27001 audit tool as both share the same security topics at the high level. The results reveal the differences in the scope of the specifications and the lack of some of the controls of KATAKRI compared to ISO/IEC 27001 and ISO/IEC 27002.

Keywords: information security management, ISO/IEC 27001, ISO/IEC 27002, KATAKRI

1. Introduction

Assuring information security is a necessity in modern organizations. There exists variation of viewpoints in information security management (ISM) concerning 'what' should be done (ISO/IEC 27000 and COBIT; IT management), 'how' it should be done (ITIL; service management), and 'who' should do it (SFIA; competence management), see (Armstrong 2013). These recommendations are used to define baseline of information security requirements ensuring that an organization has implemented the selected practices. Some of the recommendations provide the possibility for organizations to request certification, which is can then be granted if the implemented practices fulfill the audition criteria.

Widely adopted ISO/IEC 27001 prescribes a process for ISM system whereas guidance to implement security controls is defined in ISO/IEC 27002. Hence, together they comprise minimum criteria of controls and their objectives, providing also non-normative guidance for control implementation. Finnish National Security Auditing Criteria (KATAKRI) has been developed by the national authorities in Finland to verify maturity of information security practices. Approach in KATAKRI is different compared to ISO/IEC 27000 standards. As national security auditing criteria, KATAKRI defines both security control objectives and absolute security controls to meet an objective. Implementation of controls is mandatory whereas ISO/IEC 27001 leaves responsibility of the selection of controls and their implementation to organization itself by defining only the control objectives. Use of ISO/IEC 27001 is always subject to completeness of risk assessment and selection of valid security controls. On the other hand, KATAKRI may force organization to implement such controls that are not feasible from risk management or cost-benefit ratio point of view.

KATAKRI is of interest for wider than just the national audience because of its structure. It has been created in the form of the audition questionnaire, which makes it a tool that can be used in security baseline check in organizations. As information security is a process, to protect information and information infrastructure from unauthorized access, it must be defined and evaluated. ISO/IEC 27001 and 27002 specifications are not usable as audition tools themselves and hence a number of spreadsheets and applications have been created from different viewpoints to be used in the auditions. At the topic level KATAKRI could also be used as ISO/IEC 27001 audition tool, but this requires detailed analysis of the correspondence of the contents.

In our work, we study differences of security control objectives and actual controls of ISO/IEC 27001 and KATAKRI requirements to analyze completeness and mutual coverage of KATAKRI and ISO/IEC 27001. The actual comparison also takes into account ISO/IEC 27002 security control implementation guidelines, creating links between them and the security requirements in KATAKRI. First of all, however, the two specifications are

united in their terminology and structure, but whereas ISO/IEC 27002 focuses on existence of security controls to meet the security objectives, KATAKRI defines different levels of requirements that shall be fulfilled. Barlette & Fomin (2008), Fomin et al (2008), Yeniman Yildirim et al (2011), and Siponen (2006) all criticize that information security management standards focus on security process, not how well activities are carried out or how objectives are achieved. To cope with these ISMS hindrances, we create an explicit linking between a process-oriented standards and (normal) operative mode assessment in an organization.

Our analysis of KATAKRI and ISO/IEC 27002 specifications is focused to see the amount of shared common security aspects. In addition, we are interested in differences of the specifications to see the potential gaps in them, especially in the relatively new KATAKRI.

The contents of the paper are as follows: After the introduction, we provide background information on the two specifications and comparative approach in general in Section 2. Then, in Section 3 a structural comparison of specifications and high level comparison of contents of the both specifications is provided. In Section 4, we present more detailed comparison results including intersection and complements of the specifications. In Section 5 we have discussion on the results and further research.

2. Background

2.1 ISO 27000 standards

ISO/IEC 27001 is an information security standard published by the ISO/IEC standardization organization in 2005. It specifies the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining, and improving a documented Information Security Management System. ISO/IEC 27001 specifies requirements for the management of the implementation of the security controls. The detailed controls with implementation guidelines are presented in ISO/IEC 27002.

Appendix of ISO/IEC 27001 and ISO/IEC 27002 itself contain comprehensive list of controls and their objectives. Although ISO/IEC 27001 states that also additional control objectives and controls may be needed and identified from other sources. Organization defines which of the controls it shall implement. Organization may request certification against ISO/IEC 27001 for implemented ISMS. For both ISO/IEC 27001 and 27002 updated versions were released on October 2013. The changes in the update include adding of 11 new controls, but total number of the controls has been decreased from 133 to 114. Some controls have been removed and some have been merged. The number of the highest level groups of controls, security clauses, has been increased from 11 to 14.

2.2 KATAKRI – Finnish national security auditing criteria

Another approach to manage corporate security is the Finnish national security auditing criteria, KATAKRI. It is published by the Ministry of Defence, but Confederation of Finnish Industries, Finnish Communications Regulatory Authority, Ministry of Foreign Affairs, and Ministry of the Interior have also participated in the preparation of the criteria. Initial version was published in 2009 and the updated version II in 2011.

The first goal of the national security auditing criteria is to harmonize official measures while assessing organization security level. The second defined goal is "*to support companies and other organizations as well as authorities with their service providers and subcontractors to work on their own internal security*". Therefore criteria contain unofficial recommendations to help users to apply useful security practices. (KATAKRI, 2011)

2.3 Comparing standards and models

Comparing standards or methodologies may reveal several hindrances. One is the lack of widely adopted common ontology containing definitions of the basic concepts and relationships. Ramanauskaite et al. (2013) have identified that major information security management standards utilize only partially comparable security ontologies. Hence, even if standards and methodologies should lead to harmonized ontology definition, there does not exist a single widely adopted ontology definition.

Pardo et al. (2011) emphasize that in comparison it is possible to, using relationships of the models, find out how different the compared models are. Pardo et al. defines that "*in the model comparison the need to know*

the level of equality and proportion between the things being compared should take the priority". One part of comparison is terminology analysis. Pardo et al (2011) divide terminology analysis into two subtypes; syntactic analysis and semantic analysis. Our study uses only semantic analysis as the contents of the compared documents is defined in natural language and require qualitative analysis.

Multiple models can have various types of connections between them. Pardo et al. (2011) have identified four operations: union, intersection, difference, and complement. Intersection contains elements that are common in all the models and union combines together the shared contents. Difference comprises elements that the compared models do not have in common. Complement is a set of elements that are not included in one of the compared models. In this study we focus on the intersection and complements of the both specifications.

3. Structural view

3.1 Structural comparison

From structural point of view both ISO/IEC 27001 and KATAKRI controls are divided into logical groups. Following definitions are equal in both, 2005 and 2013, ISO/IEC 27002 standard versions. In ISO/IEC 27002 standard the highest level of grouping is called clauses. Each of these clauses contain "*one introductory clause introducing risk assessment and treatment*" and a number of security categories. Each security category contains one control objective and one or more controls. ISO/IEC 27002:2005 defines that control objective states what is to be achieved. The security controls in the security category can be applied to achieve the control objective. Again ISO/IEC 27002 versions 2005 and 2013 state: "*control defines the specific control statement to satisfy the control objective*". Each control is attached with the implementation guidance, which provides instructions on control implementation to meet the control objective. Definition of the implementation guidance also states that guidance may not be suitable for all organizations and other implementation options can be more appropriate. For each control there is also other information included such as references to other standards or legislation.

KATAKRI is organized as a requirements compliance questionnaire. It has four major sections called divisions, which are further divided into subdivisions. Each subdivision contains number of questions. It defines a number of requirements in the form of questions. Each question consists of a tripartite classification of requirements, corresponding to the security level concepts: the base level (level IV), the increased level (level III), and the high level (level II). These levels correspond to international security level concepts restricted, confidential, and secret, respectively. KATAKRI does not contain requirements for the highest security level, internationally known as top secret (level I).

For the KATAKRI certification the organization shall select the pursued security level. Based on selection, every requirement defined for the selected security level must be complied in each question. In addition to three security levels, there is additional set of requirements as recommendations for the industry. It contains useful security requirements recommended to all businesses to implement. For each level and industry recommendation, a number of requirements is attached. These requirements may be the same for all levels and industry recommendations, they may differ depending on the level, or higher security levels may add more requirements to the base level requirements. The questions and requirements are defined in natural language. For each question there is additional information, containing, for example, references to standards, including ISO/IEC 27002:2005, and implementation guidance.

Where KATAKRI requirements are merely ones that can be answered yes or no, ISO/IEC 27001 auditor has to evaluate that the identified set of security controls is comprehensive and implemented according to the qualitative requirements of the security controls.

ISO/IEC 27002 and KATAKRI both share the same approach grouping security concepts first on high level and then on the secondary level. In ISO/IEC 27002, highest level of grouping is division of security clauses. On the other hand, KATAKRI is divided into four divisions, which are further divided into subdivisions. Table 1 represents ISO/IEC 27002 security clause and the KATAKRI divisions and their subdivisions. ISO/IEC 27002 states that the security clauses are not in specific order concerning prioritization of the security clauses or controls. In KATAKRI prioritization is implemented in dividing security controls based on pursued security level. Hence, KATAKRI divisions and subdivisions do not relate to prioritization.

Table 1: ISO/IEC 27001 standard versions 2005 and 2013 security clauses and KATAKRI divisions and subdivisions

Logical groups of security controls		
ISO/IEC 27001:2005	ISO/IEC 27001:2013	KATAKRI
Security policy	Information security policies	Administrative security
Organization of information security	Organization of information security	Security policy, the measures guiding security action and definitions
Asset management	Human resource security	The annual security action programme
Human resources security	Asset management	Defining the goals of security
Physical and environmental security	Access control	Identifying, assessing and controlling risks
Communications and operations management	Cryptography	Security organisation and responsibilities
Access control	Physical and environmental security	Accidents, danger situations, security incidents and preventive measures
Information systems acquisition, development and maintenance	Operations security	Security documentation and its management
Information security incident management	Communications security	Security training, increasing awareness and knowhow
Business continuity management	System acquisition, development and maintenance	Reports and inspections by the management
Compliance	Supplier relationships	Personnel Security
	Information security incident management	Technical criteria
	Information security aspects of business continuity management	Securing sufficient competences
	Compliance	Other suitability of the candidate for the task
		Measures after the decision to recruit
		Measures for concluding the contract of employment
		Measures during employment
		Physical Security
		Security of area
		Structural security
		Security technical systems
		Information assurance
		Data Communications Security
		Security of Information Systems
		Security of Information
		Security of Information Handling

UML class diagram of the structures of the both documents is presented in the Figure 1. ISO 27002 standards structure is equal in both version of the standard and it contains definition of terms and their relationships. KATAKRI, on the other hand, does not contain ontology definition at all. Hence, we identified basic structures of the KATAKRI document.

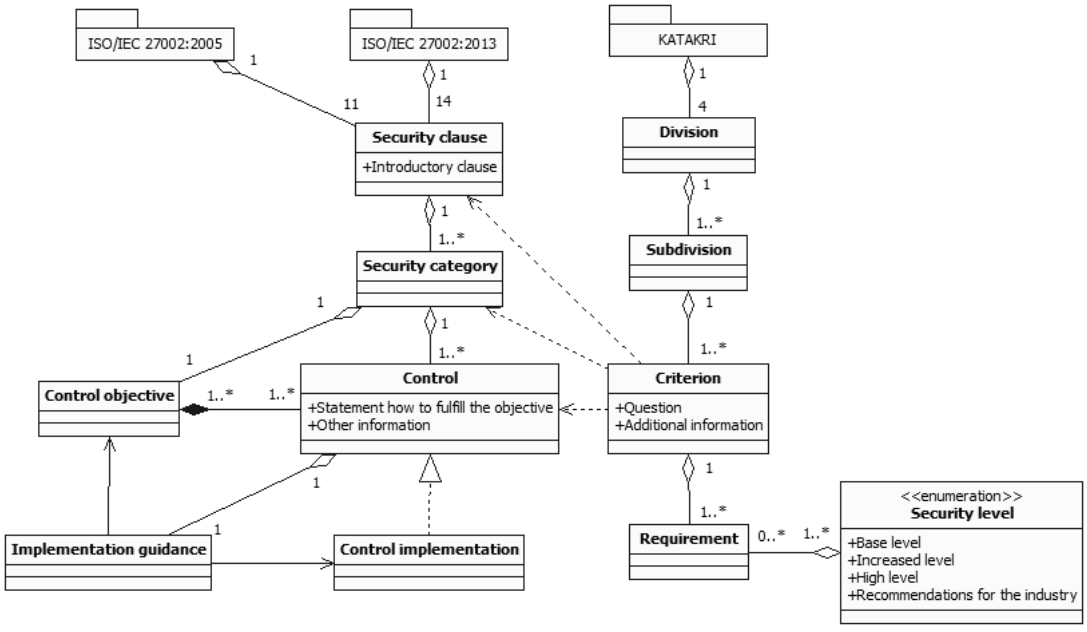

Figure 1: UML class diagram presenting structures of ISO/IEC 27002 and KATAKRI

Even if ISO/IEC 27002 and KATAKRI both share the same approach of grouping security concepts on high level, the actual structures have significant differences at lower levels. ISO/IEC 27002 standard defines control objective, which shall be achieved by implementing the defined controls. KATAKRI, on the other hand, has a question that is answered, fulfilling requirements defined for the question of the corresponding security level. Hence, KATAKRI question and ISO/IEC 27002 control objective both set goal, which is achieved by implementing defined controls or requirements.

ISO/IEC 27002 contains implementation guidance for each control that it defines. Actual implementation of the control can be done as specified in the implementation guidance or organization can select an approach that suits to its needs and characteristics (ISO/IEC 27002:2013). KATAKRI does not contain implementation guidance but provides additional information such as references to standards, legislation, and security guides.

3.2 Identified relationships

We analyzed all requirements of the KATAKRI and identified matching definitions from ISO/IEC 27002:2005. In addition we also counted number of references from KATAKRI to ISO/IEC 27002:2005. As KATAKRI defines also requirements for risk management, we included risk management requirements of ISO/IEC 27001:2005 in the analysis.

In general, the results reveal that KATAKRI had in total 432 connections to the ISO/IEC 27002:2005. From these connections 91 were direct references to ISO/IEC 27002:2005. One of these direct references is to security clause, 16 to security categories, and 74 to security controls. KATAKRI requirements had semantic equality with 21 controls. The most of the connections were semantic equality of KATAKRI requirements to implementation guidance, which we identified 320. In addition, we found out 20 connections from KATAKRI requirements to risk management section of ISO/IEC 27002:2005 and risk management requirements in ISO/IEC 27001:2005. Hence total number of identified connections was 452. Summary matrix of the connections between ISO/IEC 27002:2005 security clauses and the KATAKRI divisions is included in the appendix 0 and **Error! Reference source not found.**.

3.3 Implications of the different structures

Information security management system based on ISO/IEC 27001 and 27002 is always risk evaluation driven approach. Even though number of controls is defined in ISO/IEC 27002 specification, implementation of the controls is always a matter of evaluating suitability and appropriateness to the organization. As structurally ISO/IEC 27002 control implementation guidance provides help to find out proper control implementation, it still requires expertise from the user. The lack of the competence has been identified as one of the key obstacles to adopt ISMS by Yeniman et al (2011) and especially in small and medium sized enterprises by Barlette & Fomin (2008).

KATAKRI, on the other hand, provides more exact security requirements to be fulfilled and leaves fewer options to organization to determine appropriate way to implement the security controls. The approach of the KATAKRI may lead to situation where the requirements force organization to implement the security controls that are not feasible or have low cost-benefit ratio. Although KATAKRI requirements are more structured and specific, it does not imply that they could be neither implemented nor evaluated with lesser expertise than ISO/IEC 27002 security controls.

4. Operational view

We have divided the more specific results into four groups. First we present intersection of the two specifications. These are security controls that exist in both documents. Then we present complements of both ISO/IEC 27002 and KATAKRI, which discloses differences of the documents. More precisely, Section 4.2 contains security topics that are contained in ISO/IEC 27002 but not in KATAKRI and Section 4.3 contains the ones that are in KATAKRI but not in ISO/IEC 27002. We close the section by presenting other findings from the documents.

4.1 Intersection of specifications

In the general documents have sections that contain same topics, which can be seen as high number of links between security clauses and KATAKRI divisions as presented in Figure 2.

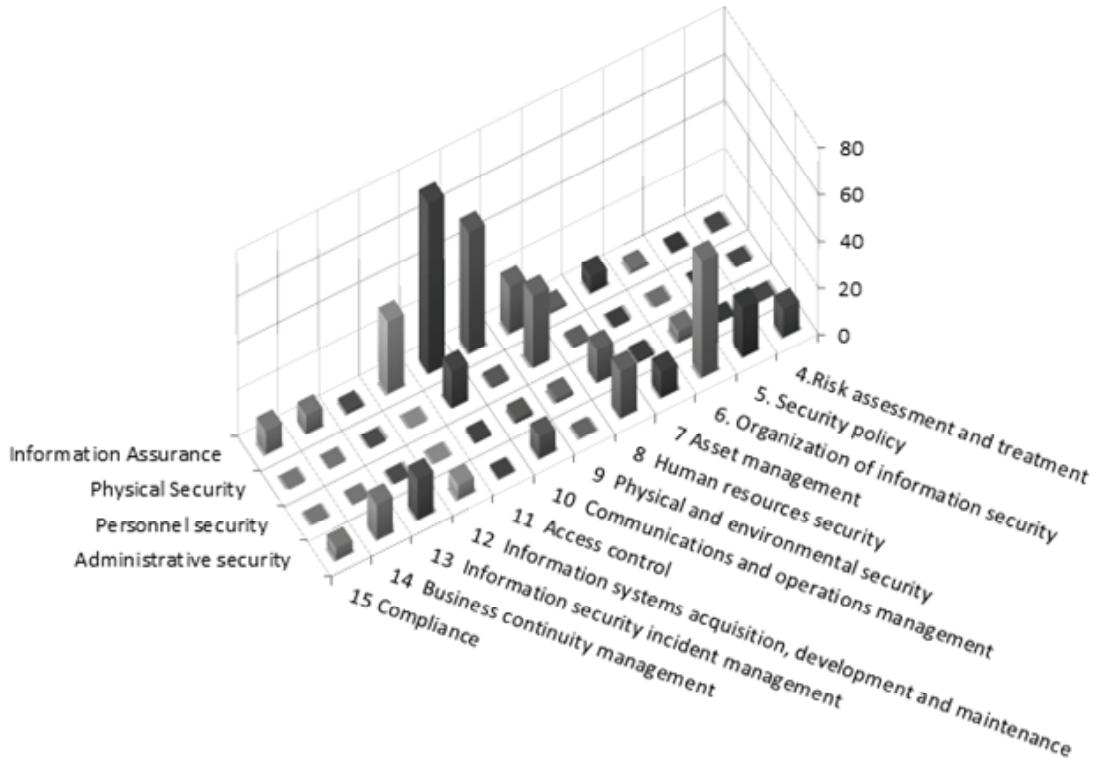

Figure 2: Number of connections between ISO/IEC 27002:2005 security clauses and KATAKRI divisions

The general security management in ISO/IEC 27002:2005 as defined in the security clauses (4-8) and (14-15) is strongly linked to KATAKRI's first division 'Administrative security'. Similarly, 'Personnel security' in KATAKRI and 'Human resource security' in ISO/IEC 27002:2005 are linked but not very strongly. Also the areas of physical security are connected. The fourth division, 'Information assurance' in KATAKRI is much dispersed related to ISO/IEC 27002:2005 covering both concrete areas in security operations (9-12) as well as higher level operations management (14-15).

In detail, several common topics that were covered by both ISO/IEC 27002 and KATAKRI were identified. Following table 2 presents intersection of the specifications divided into four domains defined by the KATAKRI.

Table 2: Intersection of ISO/IEC 27002 and KATAKRI

Common topics of information security in ISO/IEC 27002 and KATAKRI	
	Common topics
Administrative security	Security policy (22 connections) Risk management (52 connections) Security organization and responsibilities (26 connections) Incident management (8 connections) Business continuity management (32 connections)
Personnel security	Security training (36 connections) Contracts with employee (8 connections) Termination of contract (6 connections)
Physical security	Structural security (19 connections) Physical access control (26 connections)
Information security	Communication security (31 connections) Information access control (26 connections) Malware prevention and vulnerability management (12 connections) Logging (10 connections) Unauthorized devices (7 connections) Encryption (6 connections) Security of executable code (9 connections) Handling of classified information (24 connections)

Common topics of information security in ISO/IEC 27002 and KATAKRI	
	Common topics
	Systems management (10 connections) Remote work/teleworking (28 connections) Separation of production and development environments (8 connections) Backup (10 connections

The highest number of connections was in risk management as both methods require same approach to identify assets and threats to assets to perform risk mitigation. Both specifications keep security training and rising of the security awareness as an important aspect of information security.

4.2 ISO/IEC 27002 complements

We identified that KATAKRI contained in total only nine connections to ISO/IEC 27002:2005 security categories "12.1 Security requirements of information systems" and "12.2 Correct processing in applications". These two security categories contain requirements for new information system development and only nine links is relatively small amount to cover all requirements for the information system development. In the ISO/IEC 27002:2013 "12.1 Security requirements of the information systems" has been updated and category number has been changed to 14.1. Section "12.2 Correct processing in applications" and controls of it in ISO/IEC 27002:2005 have been removed from version 2013. These have been complemented with two new controls in section 14.1 of the 2013 version, but KATAKRI don't have wider correlation to either of these. Rationale for this is that KATAKRI is not meant to provide requirements for information system development, because it is audition criteria. Actually a security guideline for information system development in the state institutions, called "VAHTI 1/2013 Sovelluskehityksen tietoturvaohje", has been published. This guideline covers security requirements for the information system development. Problem has been identified also in the Finnish Defence Forces thesis by Liitsalo (2013) where she concludes that VAHTI 1/2013 has fulfilled the lack of common national guideline of generic information system development security requirements.

ISO/IEC 27002:2005 contains one security category, "10.9 Electronic commerce services", where we did not identify any links from KATAKRI. This category and contained controls have been removed from ISO/IEC 27002:2013. At the time ISO/IEC 27002:2005 was published electronic commerce was emerging and it was seen as an important domain to cover. As time passed, also many other information systems are available through the internet. Hence, electronic commerce has become only a one type among other services provided in internet, which all need to consider security in the cyber age.

ISO/IEC 27002:2013 contains controls to gather evidence in case of security incident. In KATAKRI one finds very limited requirements to cover evidence collection in case of security incidents. The KATAKRI requirements merely focus to protect audit trails, but don't include additional requirements to collect and secure the evidence.

Further complementing area in ISO/IEC 27002, compared to KATAKRI, was reporting of security weaknesses. The ISO/IEC 27002 has a specific control (13.1.1 in version 2005 and 16.1.3 in version 2013) to emphasize employee responsibility to report observed or suspected security weaknesses and vulnerabilities. KATAKRI does not contain requirement that would highlight such responsibility, even if it clearly states that for each employee the security responsibilities must be defined in their job description.

The compliance was an area where the level of details varied between specifications. Where ISO/IEC 27002 provides implementation instructions types for compliance and how to achieve compliance, KATAKRI has only the basic requirement that all operations must be compliant according to legislation.

4.3 KATAKRI complements

KATAKRI has some topics that are not part of ISO/IEC 27002 standards. On the administrative security KATAKRI contains the concept of annual security action programme, which is covered in KATAKRI subdivision A200. It is an annual plan how security will be developed comprising measures, responsibilities, schedules, and measurable results. The results of the implementation of the plan are expected to be monitored by the

management as continuous process. It is notable that there are no requirements for annual security programme at the base level, but they are included in the recommendations for the industry.

We identified number of requirements in KATAKRI that require documentation of the performed actions, but did not find equal control from ISO/IEC 27002 control objective or implementation guidance. One such topic was training, where KATAKRI requirement define that the arranged trainings must be documented, including training material and participants. ISO/IEC 27002 controls have similar control to raise awareness, but implementation guidance does not cover documentation of training. Similar widely used documentation requirement was is a job description, which is in several KATAKRI requirements referred as written definition of the responsibilities of an employee.

KATAKRI complements ISO/IEC 27002 on high security requirements. KATAKRI contains requirements that must be fulfilled to be able to handle material that is classified secret by the Finnish national definition. For the organizations that don't consider information security as competitive advantage, these controls may not be feasible to implement. These controls don't have high cost-benefit-ratio and are valid only in security critical businesses.

Hence, KATAKRI is Finnish national security audition criteria and it contains also requirements that may be illegal in other countries. Such requirements are drug tests and probationary period used in recruitment. KATAKRI also contains national requirements for physical security alarms. Such requirements are not included in the ISO/IEC 27002 standard.

4.4 Additional results

We found out also more than 20 major translation errors in KATAKRI (original version is in Finnish, which is translated to English), where a translation error caused difference in requirements. For example, in some criterions there was for certain security level "No requirements" in English version, but the original Finnish version did contain requirements.

5. Discussion

In our study we analyzed ISO/IEC 27002 versions 2005 and 2013 and compared them to Finnish security audition criteria, KATAKRI. We found out that both contain largely same security controls that security aware organizations should implement, but under a completely different structural division. Analysis also illustrates evolution of information security management trends. Results can be applied in upcoming versions of KATAKRI to evaluate the overall scope and boundaries of the security controls. They are equally relevant for ISO/IEC standardization, even if a refined version already appeared in 2013.

We identified number of common security topics that we covered by both specifications. The results reveal the different scope and lack of some of the controls of KATAKRI compared to ISO/IEC 27001 and ISO/IEC 27002. Moreover, normative controls of KATAKRI were detected, which are not included even as implementation guidance in either version of ISO/IEC 27002.

Structure of the KATAKRI makes the evaluation of the organizations information security management system easier than ISO/IEC 27002. Where KATAKRI is already structured in form of the compliance criteria, ISO/IEC 27001 requires more expertise to analyze the appropriateness of the implementation of the security controls to the specific organization. Actually the specifications complement each other well and ISO/IEC 27001 auditor may find KATAKRI as usable tool to perform security auditions. One specific difference between the two specifications is the high security level requirements contained in KATAKRI. These can be in the interest of the organizations that need to perceive very high security level in their operations. ISO/IEC 27002 implementation guidance does not contain such level of details that are included in KATAKRI's high security level requirements.

The common security topics are well covered by the both specifications and majority of controls and requirements are found in the intersection of both specifications. KATAKRI adds more specific requirements on the increased and the high security levels. Organizations having these levels of KATAKRI's security certification should be able to obtain ISO/IEC 27001 certification with little enhancements. From the structural point of view KATAKRI defines more requirements to be fulfilled and therefore organization maybe required to fulfill large number of requirements in addition to what has been acceptable in the ISO/IEC 27001 certification.

It has been noticed that SMEs have to focus more on development of their information security procedures, but most of the ISMS standards are not usable from SME organization point of view. While SMEs struggle with limited resources, but increased threads, it is important to develop new approaches that are suitable especially for SMEs. Majority of modern information security management systems, including ISO/IEC 27001, are developed for at least medium sized enterprises. One solution could be providing methods with prioritization of controls to provide, at least, a basic selection of potential roadmaps for smaller enterprises. KATAKRI contains basic prioritization using classification levels and recommendations for the industry while ISO/IEC 27002 states in the document that security controls are not in any means prioritized. Even at the lowest security level of KATAKRI, amount of controls is out reach for SMEs where security is not strategic competence area. For example, the NIST standard 800-53 (2009) defining recommended security controls for the federal information systems and organizations, contains prioritization of the security controls. Our research continues to develop methods for SMEs to enhance their security management in a cost-effective way.

Appendix 1: Total number of connections

ISO 27002:2005 and KATAKRI comparison summary (c) Riku Nykänen, 2013-2014 Total number of connections.	Administrative security	Personnel security	Physical Security	Information Assurance	
ISO 27001	8	0	0	0	8
4.Risk assessment and treatment	12	0	0	0	12
5. Security policy	21	0	0	0	21
6. Organization of information security	50	5	0	1	56
7 Asset management	11	1	0	7	19
8 Human resources security	20	14	1	1	36
9 Physical and environmental security	0	2	31	20	53
10 Communications and operations management	9	2	1	52	64
11 Access control	0	1	16	73	90
12 Information systems acquisition, development and maintenance	6	0	0	31	37
13 Information security incident management	17	0	0	1	18
14 Business continuity management	15	0	0	8	23
15 Compliance	5	0	0	10	15
Total	174	25	49	204	452

References

Armstrong, C.J. 2013, "An Approach to Visualising Information Security Knowledge" in *Information Assurance and Security Education and Training*, Springer Berlin Heidelberg, pp. 148-155.

Barlette, Y. & Fomin, V.V. 2008, "Exploring the Suitability of IS Security Management Standards for SMEs", *Hawaii International Conference on System Sciences*, Proceedings of the 41st Annual, pp. 308.

Fomin, V.V., de Vries, H.J. & Barlette, Y. 2008, "ISO/IEC 27001 information systems security management standard: exploring the reasons for low adoption", *EUROMOT 2008 Conference*, Nice, France.

ISO/IEC 27001:2005 2005, *Information technology - Security techniques - Information security management systems - Requirements*, ISO copyright office, Geneva, Switzerland.

ISO/IEC 27001:2013 2013, *Information technology - Security techniques - Information security management systems - Requirements*, ISO copyright office, Geneva, Switzerland.

ISO/IEC 27002:2005 2005, *Information technology - Security techniques - Information security management systems - Code of practice for information security management*, ISO copyright office, Geneva, Switzerland.

ISO/IEC 27002:2013 2013, *Information technology - Security techniques - Information security management systems - Code of practice for information security management*, ISO copyright office, Geneva, Switzerland.

KATAKRI 2011, *National Security Auditing Criteria version II*, Ministry of Defence, Finland.

NIST Special Publication 800-53 2009, *Recommended Security Controls for Federal Information Systems and Organizations,* National Institute of Standards and Technology.

Pardo, C., Pino, F.J., García, F., Piattini, M. & Baldassarre, M.T. 2012, "An ontology for the harmonization of multiple standards and models", *Computer Standards & Interfaces*, vol. 34, no. 1, pp. 48-59.

Ramanauskaite, S., Olifer, D., Goranin, N. & Cenys, A. 2013, "Security Ontology for Adaptive Mapping of Security Standards", *International Journal of Computers Communications & Control*, vol. 8, no. 6, pp. 878-890.

Siponen, M. 2006, "Information security standards focus on the existence of process, not its content", *Communications of the ACM*, vol. 49, no. 8, pp. 97-100.

Yeniman Yildirim, E., Akalp, G., Aytac, S. & Bayram, N. 2011, "Factors influencing information security management in small- and medium-sized enterprises: A case study from Turkey", *International Journal of Information Management*, vol. 31, no. 4, pp. 360-365.

Challenges in Information Security Protection

Teresa Pereira[1] and Henrique Santos[2]
[1]Polytechnic Institute of Viana do Castelo, Viana do Castelo, Portugal
[2]University of Minho, Guimarães, Portugal
tpereira@esce.ipvc.pt
hsantos@dsi.uminho.pt

Abstract: Security is a topic that is gaining more and more interest by organizations and government agencies. The amount of data which organizations daily have to deal with, the increasing number of on-line transactions and the lack of computer security awareness are greater motivations not only to exploit software vulnerabilities but to exploit human vulnerabilities. In general, users tend to accept new technologies with complete disregard of their security vulnerabilities, if they get sufficient benefits from them. Fostering and continuously encourage a security culture and recognizing that people still are, and will always be the weakest link, will certainly assist organizations to achieve their adequate levels of security and thus becoming closer to their business goals. Moreover, monitoring and early detection also play an important role, as it enables organizations and governmental agencies to react more quickly to events that are harder to find and understand, from the security management point of view. The rapid response to the security events and the establishment of preventive actions to manage security are starting to become a competitive strategy to organizations. In this paper we highlight some information security concepts and principles, to deliver actionable information for decision makers for managing their corporate assets and ensure their resilience.

Keywords: information security, information security management, security standards, security methodologies

1. Introduction

The proliferation of the internet-based applications has changed the way organizations conduct their businesses. Organizations are deeply interested in finding new technological initiatives at low operating cost, in order to offer better and innovative services and thus gain competitive advantage. However, with increasing reliance on technology to gain competitive advantage, information security is and has been one of the most critical and challenging requirement for conducting successful business. In fact, the new technological solutions always carry vulnerabilities, which most of time reveal unexpected security risks. In this context, organizations should identify, implement, monitoring and evaluate the most effective set of controls, to provide an adequate level of security. The security standards ISO/IEC_JTC1 (International Organization for Standardization/International Electrotechnical Commission_Joint Technical Committee and NIST (National Institute of Standards and Technology) are an important reference in the security information domain, and in general, organizations define their security program in alignment with the international security standards. However, these standards have an informative nature, lacking practical information, and relying on the interpretation of the security expert, mostly based on their experience and perception about security. The security standards ISO/IEC 27002 and the NIST 800-100 provide documented information to assist users to understand their computer security need, and thus enable them to select the appropriate controls, from a wide list of existing controls. However, these standards neither provide directives nor procedures or recommendations toward the implementation of security actions, but rather promotes a more or less arbitrary designed by practitioners in an individual basis, lacking any possibility of performance comparison. In this paper it is highlighted some information security concepts and principles, to deliver actionable information for decision makers for managing their corporate assets and ensure their resilience. This paper is structured as follows: in section 2, it is introduced the current challenges in information security; in section 3, it is presented an overview of the international security standards, and security methodology approaches; in section 4 it is presented some considerations regarding the implementation of security controls/policies; conclusions are presented in section 5.

2. Current challenges for organizational information security

Nowadays, organizations have to deal with various information security risks. Terrorist attacks, fires, floods, earthquakes and other disaster can destroy information processing facilities and critical documents. Theft and loss of organizational information can cause serious impacts on businesses' reputation, profitability, customer confidence and overall economic growth. For example, a security flaw such as the leakage of credit card information, can have negative results to a card payment companies due to the cancellation and re-issuing of compromised cards. Certainly this is costly, with high impacts on the organizations' reputation and customer

confidence. In fact, one of the fastest growing information crimes is identity theft, including customer data lost by organizations that were responsible for managing it. Such incidents have emerged from all over the world, and have resulted in the introduction of rigorous national and international data protection laws in many countries, which require organizations to protect the personal information of stakeholders. Therefore, it is very important for organizations to develop efforts to ensure their ability to securely protect their information assets and IT infrastructure. For this reason the adequate implementation of the security controls are crucial to organizations, in order to ensure their business continuity, as well as monitoring and evaluating their efficiency. Usually organizations proceed with identification and selection of the security controls according to their business needs and the associated security requirements. This security requirement should be clearly defined in the information security policy and the security policy dictate the set of controls that will provide the required protection. Additionally, to verify if the implemented controls meet the information security requirements, the evaluation and certification should be the following step. Actually, it has been noticed an increasing number of organizations seeking to obtain security certification. Certification enables organizations to comply with increasing demands from financial institutions and insurance companies for security audits. Moreover, it promotes trust in an organizations' capacity to implement appropriate security controls to manage and protect confidential client and business information. Additionally, the monitoring and evaluation of the implemented controls are crucial in order to find out if they are performing as expected, i.e., controls are in place, are designed appropriately, are operating effectively, and are monitored regularly, in an effort to reduce risk exposure. Currently, there a number of practical and deployable approaches, including a variety of globally recognized international standards and security methodologies, which can assist organizations to effectively and efficiently manage information security. In the following section will be presented an overview of these approaches.

3. Practical approaches for organizational information security management OU recognized international security standards and methodologies OU International security standards and security methodologies

Nowadays there are security standards and methodologies available to assist organizations to manage information security. To effectively create a solid structure for an information security program, organizations need to analyse and implement their methodology in alignment with either of the following international standards and methodologies.

3.1 ISO 27002 – Information technology – security techniques – code of practice for information security management

ISO 27002 is part of the ISO/IEC 27000 family of standards. It is an information security standard published by ISO/IEC, entitled Information technology – Security techniques – Code of practice for information security management. ISO/IEC 27002 is an advisory document, to define the requirements for initiating, implementing, maintaining and improving information security management practice, in order to promote confidence in inter-organizational information collaboration. The ISO 27001 replaced the BS7799-2 standard, which was primarily published as a code of practices. As this matured, a second part emerged to formally cover security management system. The ISO 27001 is harmonized with other standards. In fact, ISO 27001 is a formal specification that uses ISO/IEC 27002 to indicate suitable information security controls within the ISMS (Information Security Management System). However, since ISO/IEC 27002 is a code of practice or guideline, rather then a certification standard, organizations are free to select and implement the controls, according to their control objectives, which should reflect the organizational security requirements. Implementing ISO/IEC 27002 involves a cost-effective plan that includes appropriate security controls for mitigating identified risks and protecting the confidentiality, integrity and availability of an organization's information assets. It also involves ongoing monitoring to ensure that these controls remain effective (Saint-Germain, 2005).

The standard comprises eleven information security controls and seeks to address security compliance at all levels: managerial, organizational, legal, operational and technical. It specifies 35 control objectives, consisting of general statements of security goals for each eleven domains. The standard also includes 114 controls that identify specific means for meeting the control objectives. The ISO/IEC 27002 security domains are (ISO/IEC 27002, 2009):

- **Security policy** – demonstrate management commitment to, and support for information security.

- **Organization of information security** – develop a structure for the coordination and management of information security in the organization. Assign information security responsibility.
- **Asset management** – perform an inventory and classification of all critical or sensitive information assets.
- **Human resources security** – manage the security aspects related to employees, in order to reduce the risk of error, theft, fraud, or misuse of computer resources by promoting user training and awareness regarding the risks and threats to information.
- **Physical and environment security** – protect the information processing facilities, in order to prevent unauthorized access.
- **Communications and operations management** – manage technical security controls in systems and networks, in order to reduce the risk of failure and its consequences and develop incident response procedures.
- **Access control** – restriction of access rights to networks, systems, applications, functions and data in order to detect unauthorized activities.
- **Information systems acquisition, development and maintenance** – prevent the loss, modification, or misuse of information in operating systems and application software.
- **Information security incident management** – responding appropriately to information an security breach that was exploited by an attacker.
- **Business continuity management** – develop the organization's capacity to react rapidly to the interruption of critical activities resulting from failures, incidents, natural disasters or catastrophes.
- **Compliance** – Ensure that all laws and regulations are in conformance with information security policies, standards, laws and regulations.

3.2 NIST – National Institute of Standards and Technology

NIST is an organization from Computer Security Resources Center of the United States of America that provides several documents regarding security. Within the series of publication proposed by NIST, the 800-series encompass several aspects about computer security. Two of NIST's related documents are special publication 800-100, the Information Security Handbook: A Guide for Managers, and the special publication 800-53 - Recommended Security Controls for Federal Information Systems. NIST approach rely on providing documented information to assist users to understand their computer security needs and come up with the selection of appropriate controls and get compliant with FISMA (Federal Information Security Management Act). FISMA is United States legislation that defines a comprehensive framework to protect government information, operations and assets against natural or man-made threats. FISMA was signed into law part of the Electronic Government Act (EGA) of 2002. FISMA assigns responsibilities to various agencies to ensure the security of data in the federal government. The act requires official' programs, and the head of each agency to conduct annual reviews of information security programs, with the intent of keeping risks at or below specified acceptable levels in a cost-effective, timely and efficient manner. NIST outlines 9 steps toward compliance with FISMA (Rouse, 2013):

- Categorize the information to be protected.
- Select minimum baseline controls.
- Refine controls using a risk assessment procedure.
- Document the controls in the system security plan.
- Implement security controls in appropriate information systems.
- Assess the effectiveness of the security controls once they have been implemented.
- Determine agency-level risk to the mission or business case.
- Authorize the information system for processing.

Although FISMA is designed to federal agencies, these practices can be adapted easily to private institutions. Actually, NIST comprise a set of publications to support organizations to structure their information security programs. These documents supports security practitioner, to compare the proposed security guidelines with what currently exists in an organization and therefore help them to identify problems that demand critical

attention. The Information Security Handbook (800-100) attempts to define all of the considerations required to protect information. It treats terms such as governance, systems development life cycles, security assessments, risk management, incident response and many others in detail. The NIST 800-100 has affinities with ISO 27001/2 concerning the scope and definition of an information security program. NIST SP 800-53 - NIST Special Publications 800-53 Recommended Security Controls for Federal Information Systems and Organizations - comprises a selection of security controls for executive federal agencies (NIST 2009). The recommended security controls documented in NIST 800-53, takes NIST 800-100 to a more practical level, through the definition of the scope of the security controls, and thus the selection of the adequate controls that should be implemented. However the NIST 800-53 don't provide the context in which the controls should be applied, nor provide detailed implementation procedures for security controls, or detailed steps necessary to implement a computer security program. The lack of this document is a directive concerning what should be implemented and how. Instead, provide an extent-documented list of the existing controls, enabling to assist practitioners to understand their computer security needs and come up with the selection of appropriate controls. In fact, this is the major constraint concerning the practical application of the security standards (Pereira, 2012).

3.3 OCTAVE - operationally critical threat, asset, and vulnerability evaluation

OCTAVE is a process-driven methodology used for risk-based information security strategic assessment and planning. It helps organizations to identify, prioritize and manage information security risks.

The OCTAVE approach was developed by the Software Engineering Institute (SEI) at Carnegie Mellon University in 2001 to address the information security compliance challenges faced by the US Department of Defence.
It is intended to assist organizations to (Alberts et al*,* 2001):

- Develop qualitative risk evaluation criteria based on operational risk tolerances;
- Identify assets that are critical to the mission of the organization;
- Identify vulnerabilities and threats to the critical assets;
- Determine and evaluate potential consequences to the organization if threats are accomplished;
- Initiate corrective actions to mitigate risks and create practice-based protection strategy.

OCTAVE works in three phases:

- Phase 1: Build Asset-Based Threat Profiles
- Phase 2: Identify Infrastructure Vulnerabilities
- Phase 3: Develop Security Strategy and Plans

The overall approach of OCTAVE relies upon the creation of three catalogs of information: catalog of practices, threat profile and catalog of vulnerabilities. These catalogs then create the baseline for the organization. Regarding the evaluation process, the OCTAVE approach is formally defined in a catalog of practices, intended to assist organizations to create a protection strategy and to develop actions for the risk mitigation plans. This catalog is divided into two types of practices – strategic and operational. The strategic practices focus on organizational issues at the policy level and provide general management practices. While the operational practices focus on the technology-related issues dealing with how people use, interact with, and protect technology. Since strategic practices are based on good management practice, it is suggested that they remain stable over time. While operational practices are more subject to changes inherent to technology advances and new or updated practices should be performed in a continuously base, to support organizations to deal with those changes. In conclusion, catalog of practices intend to assist organizations to determine which specific practices are currently working well with respect to security (its current practices) and what specific weaknesses with current security practices (its organizational vulnerabilities (Alberts et al*,* 2001).

The strategic practice areas proposed by OCTAVE are following briefly described (Alberts et al*,* 2001):

- **Security Awareness and Training** – understand how information security practice is enhanced through training and education.
- **Security Strategy** – focuses on the integration of information security issues into the business strategy of the organization.

- **Security Management** – defines information security roles and responsibilities as well as management's support for information security activities.
- **Security Policies and Regulations** – addresses the organizational and management direction for information security, including the complied regulations. This area also deals with the staff's understanding of policies and enforcement of policies.
- **Collaborative Security Management** – comprise the establishment of good practices when working with third parties (contractors, Internet service providers, managed service providers, partners, etc.).
- **Contingency Planning/Disaster Recovery** – addresses plans to counteract disruptions in business activities and in systems and networks.

The operational practice areas proposed by OCTAVE are following identified (Alberts et al*,* 2001):

- **Physical Security** – comprises the physical security plans and procedures; Physical access control; monitoring and auditing physical security.
- **Information Technology Security** – comprise several areas, namely: (1) system and network management; (2) system administration tools; (3) monitoring and auditing IT security; (4) authentication and authorization; (5) vulnerability management; (6) encryption; (7) security architecture and design.
- **Staff Security** – comprise incident management and general staff practices.

OCTAVE is a comprehensive, systematic, context driven and self-directed approach, based on people knowledge. It requires the involvement of people at all levels of an organization to work together and provide inputs to identify and understand their security risks and make the right decisions about mitigation and protection. Due to its self-directed nature, the security controls implementation is highly dependent on the security experts' interpretations of the methodology adopted, and therefore, decide according to their experience and security perception about security, which might leave overlooked flaws in the organization's information security management. The essential of this approach is basically addressed by security risks analysis of the assets, and based on the assigned risk, it is defined the security requirements. In fact, the risk assessment play a crucial role in identifying potential threats to the organization and provide a perfect opportunity to implement effective controls to protect critical processes and assets. A constraint of the OCTAVE approach lies on the absence of directives for assessing and mitigating security risks. The evaluation is a process managed on the conduction of analysis workshops, to gather information and make decisions.

3.4 Other methodologies

The previous security standards and methodologies are unarguable reference in the security management domain. However there are a number of public and semi-public institutions, which provide best practices to manage security protection such as:

- Information Security Forum (ISF) (https://www.securityforum.org) provides a publication titled "Standard of Good Practice", which outlines information security best practices.
- Computer Emergency Response Team Coordination Center (CERT/CC) at Carnegie Mellon University (http://www.cert.org/). CERT/CC provides detailed and specific assistance on how to implement a security methodology.
- Information Systems Audit and Control Association (http://www.isaca.org) organize several seminars and classes of best practices.
- The International Association of Professional Security Consultants (http://www.iapsc.org/) and the Global Grid Forum (http://www.ogf.org/) provide a list of best practices.
- The portal SearchSecurity.com (http://searchsecurity.techtarget.com/) and NIST's Computer Resources Center (http://csrc.nist.gov/) are booth free portals dedicated to security which include collections of best practices.

3.5 Summing-up

This crosswalk on security standards and methodologies demonstrate some affinities according to their shared subject matter and focus, and can help organizations align adequate security practices, according to their security business requirements. A large part of effective security practice is reaching a common level of

proficiency, since patching systems in a timely way and configuring them in a secure manner increases the likelihood that an organization will remain secure. The adoption of security standards and methodologies enables organizations to build client and partner trust in their capacity to secure their information assets and ensure business continuity. The implementation of security controls, enables organizations demonstrate their commitment to secure information assets and to ensure confidentiality of customer information. They also provide their business partners and clients with greater confidence in their capacity to prevent and rapidly recover from any interruptions to production or service levels. It is becoming clear, however, that to address all aspects of security, organizations should implement a comprehensive approach using a methodological security standard. In reality, the standards are primarily intended for Governments, military and financial institutions, but they can be adopted in other contexts. NIST as well as ISO/IEC 27001 have put forth standards and guidelines intended to provide a level of protection for information resources.

Moreover, OCTAVE follows a different approach from the standards presented above. OCTVAVE proposes a catalog of practices scoped into strategic and operational context, with large involvement of organizational staff. But is this approach practicable in huge organizations? Once again, isn't security highly dependent on the peoples' perceptions about security and their interpretations of the standard or methodology adopted? Can all the procedures be applicable to all organization? Certainly not, but a lot will be, since despite the different methodologies and approaches proposed by each standard and security methodologies, they all share a lot of concepts and guidance to security protection.

4. Considerations regarding the implementation of security policies

In the previous section, it was briefly presented the most widely referenced security standards scoped in the implementation of the security good practices. It is undoubted the importance of these standards to create a solid foundation for an information security program. However, there are several enhancements needed, especially due to the changeable nature of this domain. Notwithstanding their importance, the standards are usually subject of revisions every five years, which sometimes can result inaccurate response. Moreover, despite the constraints of OCTAVE methodology, previously mentioned, it provides a more practical approach. It is driven on the asset's analysis risk, which clearly defines a realistic target-state for an information security program. The target security state will vary from one organization to another, however, in general, a target-state can be described as successfully satisfying information security requirements of critical processes and critical assets. It should be noted that establishing the target security state for any organization, regardless their business activity, is an extremely complex activity, as an organization's security expectations and requirements are often dynamic in nature. Therefore, this dynamic target-state shows more than ever, that organizations need to be agile, responsive, flexible and dynamic when they are establishing their security objectives and especially when they are conducted by an overall need for continuous improvement. In fact, ongoing risk analysis is an approach that will certainly influence the decision of what controls should be implemented to protect the critical processes and assets. Moreover, the organizational processes within a given business activity is accomplished by a specific objective which in the end is aligned to the organization's mission and objectives. In the case of critical processes, their objective is much more important as they mainly contribute towards the achievement of organizational operations. Concerning the organizational assets, they usually configure a set of items, with huge importance or value to the organizations, which may result in many forms, such as corporate policies, procedures, human resources, applications, or even information. The most critical assets from an information security perspective are those that are required by critical processes (ISO/IEC 17799:2000/270002:2005). In practice, when a business process is operating as expected, it is contributing to the organizational' objectives, and it also guarantee reasonable level of assurance, regarding its effective execution. For an organizational information security program be effective, it is required to address the analysis and resolution of information security events. Audits, statistics analysis, log analysis and performance metrics provide valuable information to assist in the evaluation and monitoring of events and potential asset's vulnerabilities. This data can then be used in the validation of a security program or even forensic analysis resulted from a security incident or critical breach. The information security incident management procedure should be explicitly documented and all employees, contractors and third-parties should be educated in its requirements and their associated responsibilities (ISO/IEC 17799:2002/27002:2005). Additionally, the analysis and measure of the effectiveness of information security also provides valuable inputs to the organization, in spite of being very complex and an extremely difficult process. In reality, it contributes to ensure that the decision managers have a detailed understanding of which assets are most valuable to them, and if those assets have been allocate to an appropriate level of classification based on

roles/responsibility, and criticality of the asset relative to prioritized level. It is recommended the documentation of the implemented policies, in a simple, standardized and structured form, but also provide the appropriate training on information security policy to all personnel within the organization, aligning their functional roles and the organization's security expectations. Almost all security standards and methodologies, law and regulations require organizations to educate their personnel, as part of the overall policy and security awareness. Fostering and continuously conduct a security culture and recognizing it as one of the single most critical core competencies within an organization, will almost certainly assist any organization in becoming closer to their business goals.

5. In conclusion

In general, to address all aspects of security, organizations first determine their primary security objectives then implement adequate methods to formalize and validate their management and finally develop procedures to achieve their objectives. In practice, a security program protection is much more than this and requires the involvement of several variables. The security standards and methodologies enable to create a solid foundation for an information security program. It is evidenced that most of the organizations analyse and implement their security programs in alignment either with the international security standards and methodologies. In reality, the adoption of security standards and methodologies enables organizations to build client and partner trust, proving their capacity to secure their information assets and ensure business continuity. For example, organizations with a committed clients and an established partner network, need to demonstrate to their partners, stakeholders, and clients that they have identified and measured their security risks and implemented a security policy and controls that will mitigate these risks. It is undoubted the importance of the security standards. However and although the security standards provide a wide list of controls, they don't give support or provide recommendations about how to perform security functions defined, as well as the procedures to perform evaluation of the implemented controls. These processes are more or less arbitrary designed by practitioners in an individual bases, lacking any possibility of performance comparison. However a clear message is passed by both standards: security is a broad and complex discipline that requires a lot of cooperation throughout the entire organization.

Acknowledgements

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: PEst-OE/EEI/UI0319/2014

References

Alberts, C., Dorofee, A., Allen, J., 2001. OCTAVE[SM] Catalog of Practices, Version 2.0. Technical Report CMU/SEI-2001-TR-020 ESC-TR-2001-020. Available from: www.cert.org/archive/pdf/01tr020.pdf [Accessed October 2011].

ISO/IEC, 2009. *ISO/IEC 2nd WD 27002 (revision) - Information technology - Security techniques – Code of practice for information security management.* ISO copyright office: Geneva, Switzerland. Available from: http://www.iso.org [Accessed September, 2010].

NIST, 2009. *SP 800-53: National Institute of Standards and Technology Special Publication 800-53, Recommended Security Controls for Federal Information Systems and Organizations* [on-line], February 2005. Available from: http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final_updated-errata_05-01-2010.pdf. [Accessed November,12 2011].

NIST, 2006. *SP 800-100: National Institute of Standards and Technology Special Publication 800-100, Information Security Handbook: A Guide for Managers. Recommendations of the National Institute of Standards and Technology*, [on-line], October 2006. Available from: http://csrc.nist.gov/publications/nistpubs/800-100/SP800-100-Mar07-2007.pdf [Accessed November, 12 2011].

Pereira, T., "Conceptual Framework to Support Information Security Risk Management". Ph.D thesis, University of Minho, 2012.

Rouse, M. (2013) "Federal Information Security Management Act (FISMA)", [on-line], TechTarget, May 2013. Available from: http://searchsecurity.techtarget.com/definition/Federal-Information-Security-Management-Act.

Saint-Germain, R. (2005) "Information Security Management Best Practice Based on ISO/IEC 17799", *The Information Management Journal,* pp 60-66.

Enrolment Time as a Requirement for Face Recognition Biometric Systems

Vítor Sá[1, 3], Sérgio Magalhães[1, 3] and Henrique Santos[2, 3]
[1]Faculty of Social Sciences, Catholic University of Portugal, Braga, Portugal
[2]Department of Information Systems, University of Minho, Guimarães, Portugal
[3]ALGORITMI Research Center, University of Minho, Braga/Guimarães, Portugal
vitor.sa@braga.ucp.pt
stmagalhaes@braga.ucp.pt
hsantos@dsi.uminho.pt

Abstract: The performance of a biometric system depends on the accuracy, the processing speed, the template size, and the time necessary for enrolment. This last factor is not much addressed in literature. In this work we collected information about the users' availability for enrolment in respect to face recognition biometrics. Were involved in testing 26 randomly chosen people. The results are presented globally, by sex, by age group and by previous experience in the use of the technology. We found that there is a generalized positive predisposition for enrolment that is expressed in some by the predisposition to try for many times and in others to try over a long time, and that it may be the youngest and the oldest the least available.

Keywords: security, biometrics, enrolment, face recognition, availability

1. Introduction

In this paper we try to understand to what extent people have patience for the process of enrolment, and argue that this is also one of the requirements of biometric systems, instantiated in this case for face recognition. This work is based on a similar study conducted for fingerprint recognition (Sá, Magalhães & Santos, 2014).

In practical implementation of biometric techniques, it is necessary to take into account the following parameters: performance - a system needs to act quickly and accurately; acceptability - people should accept the system easily; evasion - should not be easy to circumvent the system through fraudulent techniques (Singla & Sharma, 2010). Associated with the second of these parameters, methods for biometrics can also be classified as invasive or non-invasive, according to the level of nuisance that each system triggers in the user.

For the performance of a biometric system several factors contribute to it. Normally the main concern focuses on the error rate associated with authentication, leaving apart, for example, the time that is required for the enrolment process.

In the next section we present basic concepts of biometric technology, in particular of face recognition, section 3 is dedicated to the motivation and the methodology that was followed, in section 4 we describe the results obtained from the data analysis and, finally, in section 5 some conclusions are drawn.

2. Biometric technology

Biometrics is the science of measuring individual's own characteristics, making it possible the automatic recognition of people. In the context of information systems the control of who can access certain system can be made with the following methods, with its respective advantages and disadvantages: card - something an individual “has”, which can be stolen, forgotten, copied, broken, demagnetized, eventually expires, and has no cogency; password - something an individual “knows”, which can be copied, must be changed periodically and should not have personal data, and has no cogency that can causes problems in the case of forgetting or; biometrics - something an individual “is” or “does”, which does not lose validity, is not forgotten, is difficult to be copied, is true, is not transferable and is permanent.

The main components of a biometric system are the following: capture (capture of an image or basic information of biometric characteristics), extraction (through a biometric reader, geometric points are extracted, e.g., which will characterize the individual), comparison (matching with stored information) and authentication (decision about the veracity of the recognition).

Face recognition is one of the most known biometrics (Sá, Borges, Magalhães, & Santos, 2012), lying nowadays in many devices of widespread use. This technology, which reproduces computationally the natural way humans recognize people, has had a major technological development stimulated by the game industry, interested to recognize scammers in their casinos (Liu & Silverman, 2001) and, more recently, by the border control technologies. Proof of this development is the common usage that is currently being given to facial recognition algorithms incorporated both in low photographic devices, as in distributed technologies associated with social networks.

The basic concept associated with this technology is the capture of an image using a camera or video, followed by the recognition of points and/or regions that characterize a human face. These data and the relationships between them are transformed into a multidimensional vector that now constitutes the recognition pattern of the individual. Each of these steps may be performed in various ways.

The image capture phase depends primarily on the quality of the camera. However, the quality is necessarily restricted by the price factor, which determines the adoption of any technology. Associated with the price, but regardless of this, is the kind of camera that will work for normal light or via infrared. This option determines not only the capture conditions but also the processing algorithms. The quality of the camera also condition the number of colours captured, however the use of a large number of colours with implications on the computational and storage requirements may not be desirable.

After the image capture, various techniques can be applied for processing, like the image binarization, the recognition of points/regions by colour neighbourhood, the contours reduction and the subsequent segmentation. Hybrid systems can also be used (Poh & Korczak, 2001).

The above techniques are not specific in biometric recognition, to authenticate/identify the user, but are generic in computer graphics. In order to use the gathered information in the biometric recognition an intermediate stage of image cleaning is required, extracting elements subject to change, such as beard, glasses, haircut, earrings, piercings, etc.

The establishment of patterns is specific to each algorithm in order to recognize the user. Possible approaches rely on conventional statistic techniques, artificial intelligence, or both. These algorithms determine the requirements of pose and lighting (Poh & Korczak, 2001) associated with the image capture and have to take into account the aging factor.

The most important event in this area is the Facial Recognition Vendor Test (FRVT) organized by the Counterdrug Technology Development Program Office of the United States Department of Defence in collaboration with the FBI, the Canadian Passport Office, Australian Customs, the United Kingdom Biometric Work Group, among others. This event was held for the last time in 2012, now in the context of MBE - Multiple Biometrics Evaluation (NIST, 2010), with no results yet available (U.S. Department of Commerce, 2013). Already in 2006 it was achieved a False Rejection Rate of 1% for a False Acceptance Rate of 0.1% (Phillips, Scruggs, O'Toole, Flynn, Bowyer, Schott, & Sharpe, 2010). A curious aspect identified in these studies was a better performance of this technology when applied to males (Phillips, Grother, Micheals, Blackburn, Tabassi, & Bone, 2003).

The performance of biometric systems is normally associated to its accuracy, which is determined by the rate of false matches and the rate of false nonmatches (Magalhães & Santos, 2003). The first, known as FAR (False Acceptation Rate or Type II Error), measures the probability of the system to accept an unauthorized person, so the lower the probability the more reliable the system. The second, known as FRR (False Rejection Rate or Type I Error), measures the probability of the system to not recognize an authorized person, so the lower the rate the more the system will be sure of recognizing an individual. As the false acceptances decrease as the threshold increases and false rejections increase with increase of the same system requirement, there is a balance known as CER (Crossover Error Rate) or EER (Equal Error Rate), which value is used to classify a biometric system regarding its level of accuracy.

3. Motivation and methodology

In addition to accuracy to measure the performance of a biometric system, it should also be considered the following factors: the speed, which refers to how quickly a characteristic can be captured, processed into a template, and verified/identified; the size of the templates, which is the amount of bytes required to store a template; and the time necessary to the enrolment. This last factor is not much addressed in literature, so there was the reason for the study presented in this article.

During the enrolment phase, as in the recognition phase, the biometric system measures a characteristic of an individual. First it creates a digital representation of the characteristic that it wants to capture, then the digital representation is processed to create a template (a compact version of the original representation where certain features have been measured) and, finally, the template is stored internally or on an external device such as a Smart card. For this study we developed a simulator in Android environment that supposedly did these procedures.

Thus, to assess the "enrolment availability" by the user it was created an application that simulates the authentication process of face recognition. In this tool, the process appeared to fail when the user give up trying to enter his data, requesting the user to begin again the process (Fig. 1). It were recorded the number of attempts and the corresponding times. Each experiment began with a presentation, by a researcher, of the tool to the user; took place in a closed space and without the presence of any other person (even the investigator left, giving indication that would be available outside to any support); was filmed (with written consent asked to the user) with the argument that it was a scientific research, supposedly with real authentication, which would have to be documented; and terminated when the user requested the support of the investigator that, at that time, explained the true objectives of the experiment.

The study has allowed us to collect information about the users' availability for enrolment in respect of face recognition biometrics. Were involved in testing 26 randomly chosen people and mostly by academics because we assume by hypothesis that patience is distributed by people without influence of socioeconomic factors. Thus, any sample is representative for this purpose. However, for different sizes we will have different associated confidence intervals and, therefore, different error margins.

Figure 1: Simulator interface (before, during and after utilization)

4. Obtained results

We present in the following tables the synthesis of the obtained results globally, by sex, by age group and by previous experience, or not, in the use of face recognition biometric technology (Tab. 1 and 2). The age division was made according to the Sturges's Rule (Eq. 1), yielding 6 classes for 26 participants.

$$k = 1 + 3{,}3 \log n$$

Equation 1: Sturges's Rule

For the analysis of Tab. 1 we see that there is an average availability exceeding 20 attempts, and a high standard deviation in both the number of attempts as the average of the average times, which shows large differences between the behaviours of users. However, analysing the data it appears that, in general, users with less attempts are the ones who spend more time in each trial. Thus, there is a generalized positive predisposition for enrolment that is expressed in some by the predisposition to try for many times and in others to try over a long time, which reveals the existence of two psychological profiles of users as regards this biometric recognition phase.

We chose to present the data by age group despite the low representativeness of the data of each class, as the number of cases studied is relatively small (Tab. 2). However, we understand that the information have any

relevance now pointing indicators for future work. These data when divided into age classes can only be regarded as preliminary raise the possibility of being the youngest and the oldest the least available.

Still in relation to the last two columns of Tab. 1 it is apparent that the prior use of the technology of face recognition authentication does not decrease, on average, the number of attempts that the user is available to accomplish, perhaps because the registration in a face recognition system always involves repetitions of the capture process.

The results obtained were reassessed limiting our study to the first 12 trials (when they exist) of the users. In none of the studied parameters were found differences of more than one second, so it is concluded that users who have tried more than 12 times did not significantly change their behaviour over time.

Table 1: Results of the assessment of availability for enrolment

		All	Masculine	Feminine	Already used	Never used
Number of attempts	Mean	21	9	44	13	26
	Minimum	1	1	4	1	1
	Maximum	152	61	152	66	152
	Standard deviation	37.75	14.92	55.70	23.47	42.80
Minimum time		<1	<1	<1	<1	<1
Maximum time		639	639	94	355	639
Mean of mean times		86	128	8	100	62
Standard deviation of mean times		121	133	13	127	93

Table 2: Results by age group of the assessment of availability for enrolment

		a≤23	24≤a≤27	28≤a≤31	32≤a≤35	36≤a≤39	a≥40
Number of attempts	Mean	19	9	9	-	152	23
	Minimum	1	2	3	-	152	2
	Maximum	119	17	21	-	152	61
	Standard deviation	34.87	5.2	10.12	-	-	32.97
Minimum time		<1	<1	<1	-	<1	<1
Maximum time		422	639	297	-	38	572
Mean of mean times		118	30	37	-	3	140
Standard deviation of mean times		146	27	62	-	-	147

5. Conclusion

This article contains preliminary results because the sample is not large and did not address other biometrics. For example, there is the idea that people have slightly different behaviour in a biometric system by face recognition, because there is a mirror effect that will entertain the user.

The biometrics that was used in this work is the closest to others in which we have particular interest and we are looking at in terms of acceptance by the population, so the results of this work are very useful in that context. As future work we intend to do similar studies regarding other biometrics.

Acknowledgements

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: PEst-OE/EEI/UI0319/2014.

References

Jain, A.K., Flynn, P.J., & Ross, A.A. (Eds.) (2008) *Handbook of biometrics*, Springer, New York.

Liu, S., & Silverman, M. (2001). A practical guide to biometric security technology. IT Professional, 3(1), 27–32.

Magalhães, S.T., & Santos, H.D. (2003) "Biometria e autenticação", In: Associação Portuguesa de Sistemas de Informação, Porto.

NIST, (2010). MBE Multiple Biometric Evaluation, NIST, USA.

Sá, V.J., Borges, D., Magalhães, S.T., & Santos, H.D. (2012) "Biometric technologies and their perception by the common citizen", *International Journal of Electronic Security and Digital Forensics*, 4(2), 187-200.

Sá, V.J., Magalhães, S.T. & Santos, H.D. (2014) "Enrolment Time as a Requirement for Biometric Fingerprint Recognition", *International Journal of Electronic Security and Digital Forensics*, 6(1), 18-24.

Singla, S.K., & Sharma, A. (2010) "ECG as Biometric in the Automated World", *International Journal of Computer Science & Communication*, 1(2), 281–283.

Phillips, P. J., Grother, P., Micheals, R., Blackburn, D. M., Tabassi, E., & Bone, M. (2003). Face recognition vendor test 2002. In: IEEE International Workshop on Analysis and Modelling of Faces and Gestures, AMFG 2003 (p 44).

Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L., & Sharpe, M. (2010). FRVT 2006 and ICE 2006 Large-Scale Experimental Results. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(5), 831–846.

Poh, N., & Korczak, J. (2001). Hybrid Biometric Person Authentication Using Face and Voice Features. In: J. Bigun & F. Smeraldi (Eds), Audio- and Video-Based Biometric Person Authentication (Vol 2091, pp 348–353), Springer, Berlin/Heidelberg.

Retaining Control Over Private Virtual Machines Hosted by a Cloud Provider Using Mandatory Access Control, Trusted Boot and Attestation

Armin Simma and Philipp Rusch
Vorarlberg University of Applied Sciences, Austria
armin.simma@fhv.at
philipp.rusch@students.fhv.at

Abstract: The Trusted Platform Module (TPM) has been heavily criticized because it is accused of taking control over a personal computer from the owner. Interestingly and ironically this is what should be achieved - to some extent - with the system described in this paper: Control should be taken from the cloud service provider (CSP) - which is the owner of the host system with the TPM - and given to the cloud service customer. Naturally, the CSP should still be in control of its host computer system but customers should retain control over their data and virtual machines even when processed on the CSP's machine. The proposed system should enable customers to trust that their data cannot be read or manipulated by the CSP. Mandatory Access Control (MAC) prevents the CSP (including its administrator) to gain control over virtual machines processed on such a trusted system. The customer can be sure that the CSP's platform is what it claims to be. This assurance is based on Trusted Computing (TC)-technology and third party attestation that the platform of the CSP is based on a verified, tested and/or certified trusted system. This way the trust relationship between customer and CSP is not only based on personal acquaintance and contracts between customer and CSP. The trust relationship is based on both technical and organizational measures. Technical measures are MAC and TC-based technologies (TPM, integrity measurement and attestation). The organizational measure is an ecosystem extending the technical measures and consists of CSP, customer, operating system (OS) developer and a trusted third party (TTP) that tests, verifies and/or certifies trusted systems. A prototype shows that it is possible to build a system that enables cloud customers to trust that the CSP cannot access their virtual machines. Since trust should not be regarded from a technical viewpoint only, findings from social psychology and information sciences are also considered when discussing perception of trust of the cloud customer, and acceptance of cloud services.

Keywords: trust, cloud security, trusted boot, TPM, mandatory access control, attestation

1. Introduction

Cloud customers who process or store their data in the cloud want to have the same level of security as on a locally hosted system. Security comprises confidentiality and integrity of customer data. Since - in all but a handful of cases - customers cannot physically access the premises of the CSP, they cannot ascertain that the infrastructure including the host OS and virtual machine monitor of the CSP is what the CSP claims it to be. There should be a way for the CSP to prove that its infrastructure - BIOS, boot loader, host OS and virtual machine monitor (VMM) - is in a known and trusted state. Research of the last decade has produced many publications that suggest trusted boot, integrity measurement and remote attestation as a means of gaining this proof (Sailer et al, 2005; Santos et al, 2009; Schiffman et al, 2010; Neisse et al, 2011; Santos et al, 2012a; Bouchenak et al, 2013). These publications mainly focus on the technical issues on the CSP's premises to create such a trusted cloud system.

1.1 Trusted boot and integrity measurement

In the following a short overview of the technologies upon which the trusted cloud system is based is given: Trusted Computing Group (TCG) (2006) describes a process called *Trusted Boot,* which is based on *Integrity Measurement* (TCG, 2007). The basic idea is that each part of the computer system is measured before it is executed. Measuring means calculating a hash value of the binary. The measured values are stored in platform configuration registers (PCR) in the TPM. PCR values can only be written to by using the following *extend* operation: PCR = SHA1 (PCR || data) i.e. concatenate the PCR value with the data, hash the concatenation and store the result back to the PCR. Thus, a PCR value depends on all values that were extended to this PCR since the system was started.

Figure 1 shows the integrity measurement process. The measurement chain is started with the core root of trust for measurement (CRTM). The CRTM is part of the TPM and, thus, the initially trusted hardware part. The CRTM measures the BIOS code (1), saves the value in the TPM by extending a PCR (2), and then transfers execution to the BIOS. The BIOS measures the bootloader code (3), extends the measured value to the PCR (4),

after which the bootloader is executed (5). The bootloader measures the kernel code (6), extends the measured value to the PCR (7), after which the OS is executed (8). The kernel can measure application code (9), stores the value in TPM (10), and then executes the application (11). Applications can be measured and executed consecutively (12)-(14).

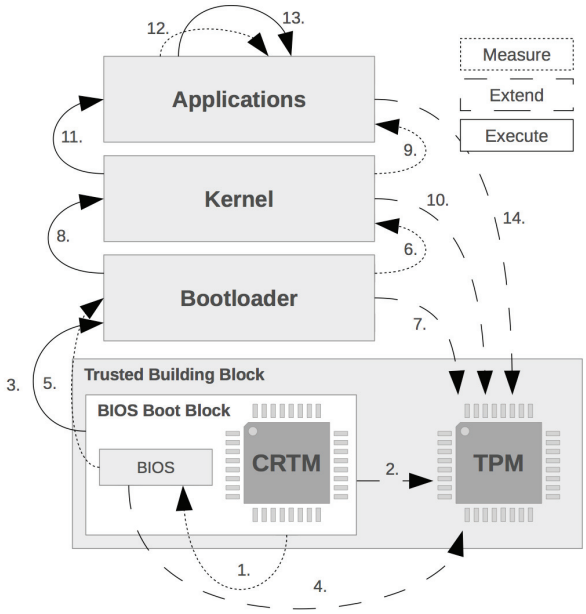

Figure 1: The measurement flow (Based on TCG, 2007)

1.2 Integrity reporting and attestation

After measurement, it must be possible to securely request the PCR values. TCG (2007) calls this process *Integrity Reporting*. Each TPM has credentials embedded. These credentials are encryption keys that are permanently embedded in the TPM by the TPM manufacturer. Integrity reporting and the credentials allow for *remote attestation*. Remote attestation allows a remote entity to request proof that a system is based on a genuine TPM and that integrity values are received from PCRs of a TPM. Remote attestation is based on digitally signing the integrity values.

Figure 2 shows the simplified process of integrity reporting: The challenger is the entity that requests the PCR values. It creates a nonce using a random number generator (RNG). This nonce is sent to the TPM. The TPM concatenates the nonce and the value of the requested PCR and hashes the result. This hash is signed using a signature key, which is generated from the Storage Root Key. The signature is sent back to the challenger. The challenger creates a hash from the concatenated nonce and reference PCR value. If the reference hash and the received hash are the same, the integrity of the PCR value is verified.

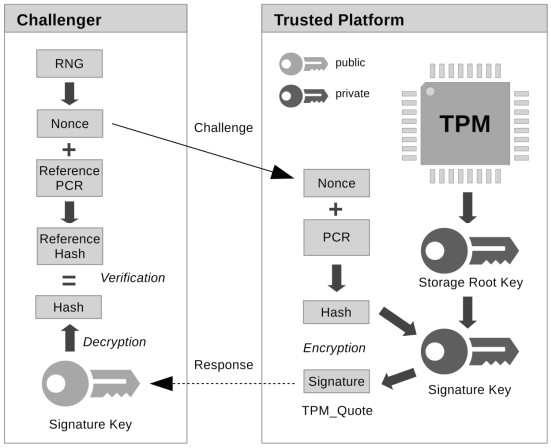

Figure 2: Integrity reporting (Based on TCG, 2007)

Using a CRTM, trusted boot, integrity reporting and attestation, the remote entity can be sure that the individual parts of the system which were executed consecutively during startup produced specific integrity values when measured.

2. Weakness and issues with trusted cloud computing based on integrity measurement and attestation

Khan, Rehman and Anwar (2011), Zhang et al (2011), Neisse, Holling and Pretschner (2011) and Selhorst, Stüble and Teerkorn (2008) have developed prototypes that perform trusted boot, integrity measurement and attestation. Different open source projects (IMA, tboot, xmhf) allow code to be downloaded and to build systems based on trusted boot, integrity measurement and attestation.

Nevertheless, such Trusted Computing (TC)-based systems could not prevail in commercial cloud implementations. Amazon AWS offers a system called CloudHSM, which - at first glance - seems to be based on a similar hardware as the TPM. The Hardware Security Module (HSM) that AWS uses is Luna SA. The aim of CloudHSM is not trusted boot but to provide a "secure key storage and cryptographic operations within a tamper-resistant hardware device". (Amazon Web Services, 2013)

To the best of our knowledge there is no commercial CSP or cloud implementation that uses TPM, TC technology and/or trusted boot. The question arises: Why does no CSP use or offer TC technology after almost a decade of research and many (big) companies supporting TC? TCG (2014) gives a list of the members of TCG.

In the following deficiencies of such TC-based systems and (possible) reasons for the reluctance of using and installing such systems in commercial or open source cloud systems are enumerated; solutions to the issues are also listed.

- *The static nature of attestation*: the difficulty of changing or updating (operating) systems that should be attested is described in Sadeghi and Stüble (2004). Each update or patch of an OS forces to create and publish new PCR values.
- *The size of the trusted computing base (TCB):* The TCB is the part of a TC-based system that must be measured before it can be executed. A huge size of the TCB leads to performance degradation of the overall system since measuring - which is based on hashing - and the TPM extend operation takes time. Experiments of Neisse, Holling and Pretschner (2011) suggest that time necessary for measuring is proportional to the size of the file. Many publications try to overcome this issue by minimizing the size of the TCB: Steinberg and Kauer (2010) built a small micro hypervisor called NOVA, which is the TCB in their system. Murray, Milos and Hand (2008) propose disaggregation to minimize the TCB. Bleikertz et al (2013) base their system on a small trusted domain builder.
- *Information leakage*: Revealing information about the host OS is not recommended. Malicious users who know details about the OS could potentially use found vulnerabilities (Santos et al 2012b). Sending PCR values to a challenger during attestation and knowing the code that created the reference values reveals all details of the host OS. Sadeghi and Stüble (2004) introduced *property-based attestation*, which allows attesting properties and not binary code. Chen et al (2006) developed a protocol for property-based attestation.
- *The complexity* of TC-based systems and - as a consequence - low acceptance by customers: It will be hard for a computer user or, in our case, a cloud customer who is not an expert in the field of IT security or TC technology to understand attestation-based and TC-based systems. If users understand all the details of a system, the level of trust in such systems usually is higher than if the systems are highly complex and the details are not or only vaguely understood. McKnight et al (2002) calls this category of trust *Situational Normality-Competence,* which is a subcategory of *Institution-Based Trust*. They showed that Situational Normality-Competence has a high influence on Institution-Based Trust. The level of trust can be increased when experts and/or highly respected organizations recommend trusting and using such systems. McKnight et al (2002, p 354) propose different strategies depending on the experience and knowledge of end users. If the end user is less experienced, they advocate that "clear explanations of structural and technological safeguards may be used to promote institutional trust". For more experienced users they propose "such mechanisms as endorsements from respected customers, seals of approval from professional associations[..]".

- *Low usability*: Usage of and trust in attestation-based cloud computing is adversely affected by low usability (Jarvenpaa, Tratinsky and Saarinen, 1999), (Heijden, Verhagen and Creemers, 2003). In the prototypical implementations mentioned above, the cloud customer has to compare the PCR values by hand. However, it would be unreasonable to ask the average customer to do so. Therefore it is proposed that additional software should run on the local computer of the cloud customer. This software is called *Cloud Attestor and Advisor* (CAA). CAA needs to be integrated with the TC-based attestation system running on the CSP's premises. CAA compares the PCR values received during the attestation process with the integrity values published by a trusted third party. CAA software is what the customer directly interacts with. Therefore trust level of customers into CAA must be very high.
- *(Missing?) Recommendations and reputation:* In the following CAA is compared with antivirus software since both perform security-relevant tasks, both download comparative data from a remote entity – called signature files in antivirus software - and both must be trusted by end users if end users should accept them. Antivirus and anti-malware software is well established. A study of McAfee (2012) shows that 83% of participants of the study have working security protection, which includes antivirus software. 39% of the participants in a study from McAfee (2011) answered they use free software for security protection purpose. Many computer users do not care or do not have the knowledge to care about the details of any found malware. Gross J. and Rosson M. (2007) discovered in a survey that 58% cannot distinguish between virus and worm and 25% of them are not even aware they have a virus scanner on their system. The antivirus software installed on the local system receives malware signatures from a central malware signature database. This signature database (as well as the software) comes from a trusted entity. The same should be implemented in TC-based and attestation-based cloud ecosystems: trusted reference values for PCRs are downloaded from a separate trusted entity and the local (trusted) software handles the necessary attestation and integrity checks. It should be emphasized at this point that most users trust antivirus software without detailed knowledge about the technologies behind it. The following question arises: Where does the high trust level result from? The hypothesis of the authors is that the contributing factors are recommendations by experts and/or the media, usability and the reputation of the companies developing the protection software.

3. The five pillars for trusted clouds

The authors propose an ecosystem for Infrastructure as a Service (IaaS) cloud that extends the technical controls (trusted boot, measurement and attestation, MAC) with the help of a trusted third party and an (indirect) certification and audit system: A trusted third party certifies that the software (including bootloader, OS and VMM) the CSP is using conforms to specified security requirements. The TTP has the task to test and check that the BIOS, bootloader, host OS and the virtual machine monitor conform to specified requirements defined in advance. The TTP does not have to check this on the physical premise of the CSP. The test, check and audit of the system can be done locally.

The attestation that the CSP is using the specific certified system is done using trusted boot, a chain of trust, integrity measurement and remote attestation. To confine the root user, a MAC system is integrated in the CSP system.

The secure and trusted ecosystem should be based on five pillars; the first four are technical measures, while the last pertains to the organizational architecture:

1. Hardware root of trust: Our system is based on the TPM. This part of our ecosystem is specified in the TPM main specification of TCG (2011). It should be possible to use any hardware root of trust that has the possibility to measure upper layer firmware or software and is able to store the measurements in a secure (tamper-proof) way. The CSP and the consumer must trust the hardware root of trust.
2. A complete chain of trust during trusted boot: It is essential for the overall security and integrity of the ecosystem that the chain of trust has no gaps between the individual blocks that are measured and executed. This has been described by Kauer (2007) and Martin (2008). Trusted boot has been introduced in section 1.1 and is described in TCG (2006).
3. Integrity Reporting and Attestation: This has been introduced in section 1.2. Details can be found in TCG (2006) and TCG (2007)
4. MAC to confine users including the root user: Regarding possible adversaries against a cloud system, it has been discovered that insider attackers are among the nine most critical threats to cloud systems (Cloud

Security Alliance, 2013). Since administrators (root users in Linux systems) usually have all access rights, the root users are the most powerful insiders and thus - if they are malicious - the most critical insider attackers. Thus, a trusted cloud ecosystem must ensure that a malicious root user cannot harm the system, steal data or engage in any other malicious activity. Shashidharan and Jitesh (2011) suggest using a MAC system to confine the root user. Nahari (2007) describes a system that combines MAC, trusted boot and embedded Linux. Sailer et al (2005) integrated a MAC system into a XEN hypervisor. Bleikertz et al (2013) consider five actors in their model as being possible adversaries, one of them being the root user. Their protection of the cloud against root users is based on MAC for low-level resources and on a de-privileged Dom0. In a typical Xen implementation Dom0 is the domain that allows direct interaction with the hypervisor. Dom0 allows for starting, stopping or managing other domains.
5. Trust Ecosystem: The author's contribution will focus on the ecosystem surrounding the technological base on the CSP's premises. The authors advocate separating the entity that assumes the role of the challenger in figure 2 and the entity that reviews the code used for the platform.

The former is software on the cloud customer site (CAA) and deserves closer attention since this is what the customer interacts with; the latter is a TTP that reviews code, creates digital signatures of the binaries and publishes these reference integrity values. In the following section the complete trust ecosystem is described in detail.

4. The trust ecosystem

To explain the two entities - introduced with the fifth pillar in section 3 - the ecosystem is compared to code signing combined with public key infrastructure (PKI). In a PKI a digital certificate signed by a certificate authority – the TTP - attests authenticity of the public key of the certificate holder. A software signature is an encrypted hash of code. Signing is done with the private key of the signer. The corresponding public key is authenticated with the certificate.

The similarities between our trust ecosystem and PKI are that both use code signing and both require a TTP. The main difference is the category of trust that is provided: PKI provides *identity trust;* our system should provide *provision trust.*

PKIs do neither guarantee anything about the reliability of the certificate holder nor state anything about the quality of service provided by the certificate holder. But the certificate increases the level of trust. Josang et al (2007) describes categories of trust, two of them being *identity trust* and *provision trust.*

In our trust system identity trust is a prior condition but provision trust is postulated as well: The cloud customer should be enabled to trust that the platform of the CSP is reliable and that confidentiality and integrity of the customer's data is assured. To achieve provision trust in our cloud ecosystem an additional entity is introduced. The term Trusted Third Tester (TTT) is used for this entity because it is more than a TTP in a PKI, which provides only for identity trust. The TTT checks, tests, reviews and/or audits the software platform that is used by the CSP. If the trust level in TTT is high, trust in CSP can be improved if technical pillars 1 to 4 from section 3 are implemented and used in the ecosystem.

Figure 3 gives an overview of our proposed ecosystem: (1) The developer develops the OS, the bootloader or the firmware. Examples for this entity are the open source community or a commercial software company. If it is the open source community there must be a coordinating organization that requests certification and code review from the trusted third party. (2) The TTT tests, verifies and audits the software. The TTT can be the open source community, a commercial or a (semi-)governmental organization. (3) The TTT signs and publishes the integrity values. (4) On the premises of the CSP the specific VMM and/or host OS is running. Each physical host of the CSP is equipped with a TPM. (5) and (6) comprise the integrity reporting process from figure 2. CAA software is the challenger, the TPM responds with signed PCR values. (7) CAA compares the PCR values received from TPM with the reference values received from TTT in (3).

In contrast to the entities from figure 3 the entities enumerated in TCG (2006) are not specific to cloud systems but cover general trust infrastructures.

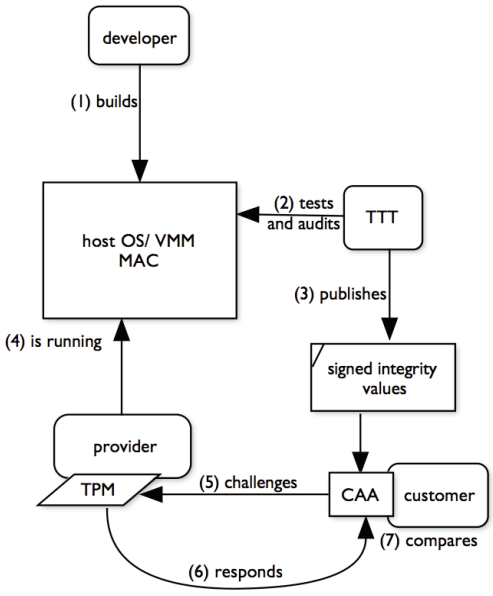

Figure 3: The trust ecosystem

5. Implementation

The authors are working on a prototype of the overall trust ecosystem. Up to now the software that runs on the CSP's site is implemented. The prototype is implemented using open source software: It is based on a Linux system (Debian 7.2.0) including SELinux. Integrity measurement architecture (IMA) and software for TPM usage allow for trusted boot and attestation. Figure 4 gives an overview of the architecture.

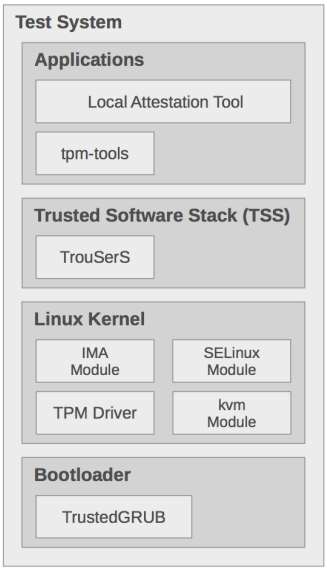

Figure 4: Architectural overview of the prototype

In the following the building blocks of figure 4 are listed and described. The software is either part of the Debian distribution or it can be downloaded from sourceforge.net with the exception of the Local Attestation Tool.

- TrustedGRUB is an extension of the GRUB bootloader to allow for trusted boot.
- IMA module is a kernel module that allows for integrity measurement. It was first presented by Sailer et al (2004).
- TPM Driver is a kernel device driver for Linux that enables the TPM.
- SELinux is a MAC system.
- KVM is a full virtualization solution for Linux.

- TrouSerS is a library to access the TPM functionality.
- tpm-tools comes with TrouSerS. It contains the commands to interact with the TPM.
- Local Attestation Tool is a client-server application developed by the authors that allows for integrity reporting as described in section 1.2.

6. Conclusion and further research

Missing trust in CSPs is one of the key inhibitors for further acceptance and propagation of cloud technologies (Gupta et al, 2013). TC is one of the most promising technologies to establish trust in cloud systems. Furthermore, cloud systems are one of the most appealing applications of TC. To promote TC in cloud systems, further research is needed that combines technical solutions with findings from behavioral science, social psychology and usability research.

Many publications and surveys exist that analyze intentions and trust of users of web stores, web services or Internet auctions. There are also publications that analyze trust in cloud systems (Uusitalo et al, 2010; Pearson and Benameur, 2010; Pearson, 2013). However, further research is necessary to be able to transfer these findings to the field of TC-based cloud computing.

The reputation of web service providers is an influencing factor for the level of trust (Jarvenpaa, Tratinsky and Saarinen, 1999). Assuming that online stores and cloud platforms basically have the same trust factors, the following actions are needed: Organizations (TTTs) able to perform diligent tests and reviews of source code of cloud platforms should be established. Trust in these organizations is fundamental. To increase the reputation of – and as a consequence trust in - these organizations, experts from the scientific community or governmental organizations should promote and recommend TTTs.

The open source community could also be a TTT. It is predestined to perform reviews of open source code for cloud systems. Reviews can be performed decentralized but the creation, signature and publication of integrity values should be done by one central organization, which has to be trusted by customers. One key advantage of the open source community being TTT is the fact that it is not mistrusted as some governmental organizations currently are as shown by a survey of Pew Research Center (2013).

Heijden, Verhagen and Creemers (2003) state that there is a strong positive correlation between the size of an organization and its level of trust. Separating the developer, CSP and TTT allows small-sized CSPs – which usually are less known and thus less trusted – to become “certified” by TTT. TTT attests that this CSP uses a trusted platform, thereby increasing the trust level into the CSP.

Our ecosystem provides no solution to the problem of information leakage when using binary attestation. Property-based attestation should be integrated with the ecosystem. Kostiainen, Asokan and Ekberg (2011) enumerate deficiencies of property-based attestation, one of which is the required existence of a TTP. Our ecosystem already contains such a TTP.

A factor that could increase trust in the proposed TC-based system is usability. Karppinen (2012) states that “ease of use has the biggest effect” on trusting cloud services. In the discussion section of Roy et al (2001) it is stressed that certain usability criteria are key factors for high trust levels. Further research and action is needed to improve user experience.

In our prototype attestation is done using the Local Attestation Tool. This should be extended to allow for remote attestation. In section 2 it was stressed that such CAA software is what the customer directly interacts with; therefore special focus must be put on the usability of CAA.

TC-based cloud technologies are adversely affected by mistrust in TPM. One of the reasons for this mistrust is that TPM often is used synonymously with digital rights management (Schneier, 2005; Law, 2006). To counter such arguments clear explanations of the technologies are essential. It should be emphasized that TPM - when used for remote attestation - does not prevent the usage of any software product or vendor. It is the cloud customer who decides whether or not to send his data or virtual machine to the cloud.

References

Amazon Web Services (2013) "AWS CloudHSM Product Details " [online], http://aws.amazon.com/cloudhsm/details/ [Accessed 30 January 2014]

Bleikertz, S. et al (2013), "Client-Controlled Cryptography-as-a-Service in the Cloud", in Jacobson, M., Locasto, M., Mohassel, P. and Safavi-Naini, R. (Eds.), *Applied Cryptography and Network Security*, Lecture Notes in Computer Science, Springer Berlin Heidelberg, Vol. 7954, pp. 19–36.

Bouchenak, S. et al (2013), "Verifying cloud services: present and future", *ACM SIGOPS Operating Systems Review*, Vol. 47 No. 2, pp. 6–19.

Chen, L. et al (2006), "A protocol for property-based attestation", *Proceedings of the first ACM workshop on Scalable trusted computing*, ACM Press, pp. 7–16.

Cloud Security Alliance (2013) "The Notorious Nine. Cloud Computing Top Threats in 2013" [online], https://cloudsecurityalliance.org/download/the-notorious-nine-cloud-computing-top-threats-in-2013/ [Accessed 30 January 2014]

Gross, J.B. and Rosson, M.B. (2007), "Looking for trouble: understanding end-user security management", *Proceedings of the 2007 symposium on Computer human interaction for the management of information technology*, ACM Press, p. 10.

Gupta, S., Kumar, P. and Abraham, A. (2013), "Cloud Computing: Trust Issues, Challenges, and Solutions", *Managing Trust in Cyberspace*, Taylor & Francis, pp. 13-40

Van der Heijden, H., Verhagen, T. and Creemers, M. (2003), "Understanding online purchase intentions: contributions from technology and trust perspectives", *European Journal of Information Systems*, Vol. 12 No. 1, pp. 41–48.

Jansen, B., Ramasamy, H.-G.V. and Schunter, M. (2008), "Policy enforcement and compliance proofs for Xen virtual machines", *Proceedings of the fourth ACM SIGPLAN/SIGOPS international conference on Virtual execution environments*, ACM Press, pp. 101–110.

Jarvenpaa, S.L., Tractinsky, N. and Saarinen, L. (2006), "Consumer Trust in an Internet Store: A Cross-Cultural Validation", *Journal of Computer-Mediated Communication*, Vol. 5 No. 2, pp. 1–5.

Jøsang, A., Ismail, R. and Boyd, C. (2007), "A survey of trust and reputation systems for online service provision", *Decision Support Systems*, Vol. 43 No. 2, pp. 618–644.

Karppinen, K. (2012), "Users' Trust and Secure Feeling towards Cloud Services", *Proceedings of the 2012 International Conference on Advances in Computer-Human Interactions*, pp. 360–365.

Kauer, B. (2007), "OSLO: Improving the Security of Trusted Computing", *Proceedings of 16th USENIX Security Symposium*, USENIX Association, Berkeley, CA, USA, pp. 16:1–16:9.

Khan, I., Rehman, H. and Anwar, Z. (2011), "Design and Deployment of a Trusted Eucalyptus Cloud", *Proceedings of the 2011 IEEE 4th International Conference on Cloud Computing*, IEEE, pp. 380–387.

Kostiainen, K., Asokan, N. and Ekberg, J.-E. (2011), "Practical Property-Based Attestation on Mobile Devices", in McCune, J., Balacheff, B., Perrig, A., Sadeghi, A.-R., Sasse, A. and Beres, Y. (Eds.),*Trust and Trustworthy Computing*, Lecture Notes in Computer Science, Springer Berlin Heidelberg, Vol. 6740, pp. 78–92.

Law, S. (2006) "The Trusted Platform Module" [online], https://chillingeffects.org/anticircumvention/weather.cgi?WeatherID=534 [Accessed 30 January 2014]

Martin, A. (2008), *The Ten Page Introduction to Trusted Computing*, Tech. Rep. RR-08-11, Oxford University Computing Laboratory.

McAfee (2011) "McAfee Reveals Average Internet User Has More Than $37,000 in Underprotected 'Digital Assets'" [online], www.mcafee.com/au/about/news/2011/q3/20110927-01.aspx [Accessed 30 January 2014]

McAfee (2012) "Consumer Alert: McAfee Finds One in Every Six Personal Computers Have Zero Protection", [online], www.mcafee.com/cn/about/news/2012/q2/20120530-01.aspx [Accessed 30 January 2014]

McKnight, D.H., Choudhury, V. and Kacmar, C. (2002), "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology", *Information Systems Research*, Vol. 13 No. 3, pp. 334–359.

Murray, D.G., Milos, G. and Hand, S. (2008), "Improving Xen security through disaggregation", *Proceedings of the fourth ACM SIGPLAN/SIGOPS international conference on Virtual execution environments*, ACM Press, Seattle, WA, USA, pp. 151–160.

Nahari, H. (2007), "Trusted Secure Embedded Linux", *Proceedings of the Linux Symposium*, Ottawa, Canada, pp. 79–86.

Neisse, R., Holling, D. and Pretschner, A. (2011), "Implementing Trust in Cloud Infrastructures", Presented at the 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, IEEE, Newport Beach, CA, USA, pp. 524–533.

Pearson, S. (2013), "Privacy, Security and Trust in Cloud Computing", in Pearson, S. and Yee, G. (Eds.),*Privacy and Security for Cloud Computing*, Computer Communications and Networks, Springer London, pp. 3–42.

Pearson, S. and Benameur, A. (2010), "Privacy, Security and Trust Issues Arising from Cloud Computing", Presented at the 2010 IEEE Second International Conference on Cloud Computing Technology and Science, pp. 693–702.

Pew Research Center (2013) "Trust in Government Nears Record Low, But Most Federal Agencies Are Viewed Favorably" [online], www.people-press.org/2013/10/18/ [Accessed 30 January 2014]

Roy, M.C., Dewit, O. and Aubert, B.A. (2001), "The impact of interface usability on trust in Web retailers", *Internet Research*, Vol. 11 No. 5, pp. 388–398.

Sadeghi, A.-R. and Stüble, C. (2005), “Property-based attestation for computing platforms: caring about properties, not mechanisms”, *Proceedings of the 2004 workshop on New security paradigms*, ACM Press, pp. 67–77.

Sailer, R. et al (2005), “Building a MAC-Based Security Architecture for the Xen Open-Source Hypervisor”, Presented at the Annual Computer Security Applications Conference, IEEE, pp. 276–285.

Sailer, R. et al (2004), “Design and Implementation of a TCG-based Integrity Measurement Architecture”, *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, SSYM’04, USENIX Association, Berkeley, CA, USA, pp. 16–16.

Santos, N., Gummadi, K.P. and Rodrigues, R. (2009), “Towards Trusted Cloud Computing”, *Proceedings of the 2009 conference on Hot topics in cloud computing*, Article No. 3.

Santos, N., Rodrigues, R. and Ford, B. (2012a), “Enhancing the OS Against Security Threats in System Administration”, *Proceedings of the 13th International Middleware Conference*, Middleware ’12, Springer-Verlag New York, Inc., New York, NY, USA, pp. 415–435.

Santos, N. et al (2012b), “Policy-sealed Data: A New Abstraction for Building Trusted Cloud Services”, *Proceedings of the 21st USENIX Conference on Security Symposium, Security’12*, USENIX Association, Berkeley, CA, USA, pp. 10–10.

Schiffman, J. et al (2010), “Seeding clouds with trust anchors”, *Proceedings of the 2010 ACM workshop on Cloud computing security workshop*, ACM Press, pp. 43–46.

Schneier, B. (2005) "Trusted Computing Best Practices", [online], https://www.schneier.com/blog/archives/2005/08/trusted_computi.html [Accessed 30 January 2014]

Selhorst, M., Stüble, C. and Teerkorn, F. (2008), “Introduction and Analysis of the Open Source TCG Software Stack TrouSerS and Tools in its Environment”, Study on behalf of the German Federal Office for Information Security (BSI), Bonn (Germany), Sirrix AG

Shashidharan, A. and Shah, J. (2011) "Enabling Cloud Customers to Trust the Cloud", [online] http://jiteshs.com/cloud-trust.pdf [Accessed 30 January 2014]

Steinberg, U. and Kauer, B. (2010), “NOVA: a microhypervisor-based secure virtualization architecture”, *Proceedings of the 5th European conference on Computer systems*, ACM Press, pp. 209–222.

TCG (2006) "TCG Infrastructure Work Group Architecture Part II - Integrity Management", Specification Version 1.0, Rev.1.0 [online] https://www.trustedcomputinggroup.org/resources/infrastructure_work_group_architecture_part_ii__integrity_management_version_10 [Accessed 28 January 2014]

TCG (2007) "TCG specification architecture overview", Rev.1.4 [online] www.trustedcomputinggroup.org/files/resource_files/AC652DE1-1D09-3519-ADA026A0C05CFAC2/TCG_1_4_Architecture_Overview.pdf [Accessed 28 January 2014]

TCG (2011) "TPM Main Specification", Specification Version 1.2, Rev.116 [online] https://www.trustedcomputinggroup.org/resources/tpm_main_specification [Accessed 28 January 2014]

TCG (2014) "TCG Members" [online], www.trustedcomputinggroup.org/about_tcg/tcg_members [Accessed 28 January 2014]

Uusitalo, I. et al (2010), “Trust and Cloud Services - An Interview Study”, Presented at the IEEE Second International Conference on Cloud Computing Technology and Science, IEEE, pp. 712–720.

Zhang, F. et al (2011), “CloudVisor: retrofitting protection of virtual machines in multi-tenant cloud with nested virtualization”, *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles*, ACM Press, pp. 203–216.

Cyber Security and Civil Engagement: Case of Lithuanian Virtual Community Projects

Aelita Skaržauskienė, Agnė Tvaronavičienė and Gintarė Paražinskaitė
Mykolas Romeris University, Vilnius, Lithuania
aelita@mruni.eu
a.tvaronavciene@mruni.eu
giparaz@mruni.eu

Abstract: Lithuania, like other European countries is constantly faced with low civil engagement, for example, involvement in decision making processes, the electoral process, volunteering, etc. Despite the various policies and strategies of European Union and the state level to grant the priority of interactions of governance and inhabitants, the situation remains the same. In order to reveal the main reasons for general low participation, many scientific researches covering different aspects were done. This article focuses on the analysis of the legal concerns constantly raised by scientists and respondents of public opinion survey on what restricts the broad involvement of virtual communities in civil activities. As a result, the main legal risks in this context in Lithuania are identified. For this systematic and comparative analysis, generalization, document analysis were completed and combined with results of quantitative research from Lithuania. Aspects such as privacy, security of personal data, demand for identification, internet censorship are discussed. The results of the research prove that security is among the top 5 reasons that prevent participation in virtual communities. While active users of virtual communication are mostly worried about threats for personal data, for intellectual property and violation of rights and obligations. The topic is related to cyber safety as it identifies the legal risks for the members of virtual communities acting in Internet and must be considered as relevant to the issues of cyber intelligence and security. The value/novelty of the article is to identify the legal threats associated with using such social technologies as virtual communities for promoting the inclusion of society into the decision-making process, taking into account that existing situation and public opinion in Lithuania is not yet analyzed. The methodology and results might be used as a case or part of comparative analysis in other European countries that face the same problems.

Keywords: social technologies, virtual community, collective intelligence, legal issues, privacy, censorship

1. Introduction

Cyber security issues, within the context of civil engagement, are important aspects, which must be analysed and discussed in the field of science and practice. Various social technologies have created possibilities for society members to communicate despite the physical distance, therefore the new opportunities have brought high prospects for more sophisticate crimes and other violations of rights and obligations of users, administrators and states, as regulatory bodies. This article connects several independent fields of research: social technologies, civil engagement, collective intelligence and cyber security in order to reveal the main threats of using social technologies during the process of engaging society into socially responsible activities.

Most people today do not imagine their life without social technologies. We are used to sharing information, communicate, pay taxes, host contests, buy and sell goods on web. Executive Director of Google, Eric Schmidt has stated that technologies are no longer about hardware or software, but “about mining and use of enormous data to make the world a better place” (Morozov, 2013). So the big part of everyday life routine has already been transferred from physical reality to virtual. As the second level of such transformation, is directing new technologies towards creation of greater welfare. Face-to-face interaction is not fading away nor there is increasing isolation of people in front of their computers (Castells, Cardoso 2005). On the contrary, Internet users are more social, have more friends and contacts, and expected to be more socially and politically active than non-users. Especially as the anticipation of the younger groups of the population substantially increase sociability by the new forms of wireless communication (e.g. mobile phone). Communication on the Internet by means of Web 2.0 and social media tools has develop new forms of collaboration, group knowledge, on-line Collective Intelligence (Malone, 2010, Salminen, 2012), social networking (Gunawardena et al., 2005). The process of creating innovative technologies socially oriented goals has become one of the competitive advantages for organisation and communities. This tendency is sensed even from Silicon Valley motto change (from “innovate or die” to “ameliorate or die”) where emphasis is on abilities to change things and to get humans to behave in more responsible and sustainable ways in order to maximise efficiency (Morozov, 2013). Previous documented attempts to create better opportunities for greater welfare, and other innovations, are accompanied with many risks. While people act on the web, they

must also preserve their personal data; refrain from certain actions, which may violate rights of other people and so on. Such risks form a big part of the research field of cyber security, which leads to the necessity to explore this question broader and in an interdisciplinary manner.

The present article focuses on the legal issues of cyber security that were identified during the empirical research on the involvement process of virtual communities in different civil activities. Aspects such as privacy, security of personal data, demand for identification, Internet censorship are discussed. The purpose of the article is to analyse the legal concerns, often raised by scientists and respondents of public opinion survey, held in 2013 in Lithuania, on what restricts the broad involvement of virtual communities in civil activities. As a result the main legal risks in this context in Lithuania are identified. For this systematic and comparative analysis, generalization, document analysis were done and combined with results of quantitative research (social opinion survey) from Lithuania.

The legal threats of using social technologies as virtual communities to promote the inclusion of society into decision making, especially with focus on collective intelligence, taking into account existing situation and public opinion in Lithuania, were not yet analysed. The methodology and results might be used as a case or part of comparative analysis in other European countries that face the same problems.

2. Background of the research

2.1 Main concepts of the research: collective intelligence, civil engagement and cyber security

One field where application of innovative technologies builds bridges for new quality of social decision making is activity of virtual communities, which leads to the emergence of collective intelligence. The concept of collective intelligence may be revealed by taking into consideration the few main characteristics: a) it is an activity for a group of people; b) it is an activity which is oriented to create or decide something; c) the results of such activity is more effective than individually proposed decisions (Tvaronavičienė, Paražinskaitė, 2013). In brief, collective intelligence may be defined as the collective, intellectual activity of a group of individuals (Malone, Laubacher, Dellarocas, 2009). Collective intelligence, which emerged in the activities of virtual communities, is a new quality of civil engagement that grants more effectiveness and compliance with societal needs. Civil engagement refers to the ways in which citizens participate in the life of a community in order to improve conditions for others or to help shape the community's future (Adler, Goggin, 2013). The interaction between social technologies and civil engagement creates an environment, where collective intelligence has all opportunities to emerge and be used. Actually, it is one of the most modern examples of social technologies working towards a better future. Of course good intentions of the generators usually build high opportunities, thus the realization and active usage of social-oriented production, as well as other ones in virtual space, have many risks. Most of them are the issues of cyber security, which in a broader sense may be understood as protection against disruption and misuse of Internet facilities (Gradi, Parisi, 2006). Risks and responsibilities in the hyper connected world are among number one topics in the world. The initiatives to make institutions more cyber resilient and creating models for protection from cyber-attacks are becoming less effective despite the efforts (*Chinn, Kaplan, Weinberg 2014). In c*oncluding the context of this article, it is necessary to state that social technologies as a medium between people and bodies to make decisions, has not only created comfortable platforms for spreading the ideas, but also presented opportunities to evaluate and improve individual ideas and convert them into collective intellectual productions, which due to synergistic effects has undoubtedly new quality and applicability. Thus for broad application of it, attention must be paid towards cyber security issues in order to develop a safe and reliable environment for people, who wish to engage and generate ideas for the greater welfare of society.

2.2 The current state of civil engagement in Lithuania

Lithuania, like other Central and East European countries, is often faced with low civil engagement. According to the results of public opinion survey (Vilmorus, 2013) majority of respondents (Lithuania residents) are not active in participating in activities connected with civil engagement. For example during the last five years, only 12 per cent of respondents have signed petitions or have taken part in commenting on the web some social, political or economic questions, only 9 per cent of respondents were involved in public discussions as well as only 4 to 8 per cent of respondents were involved in other civil activities (protests, demonstrations and etc.). Taking in consideration the participation in activities of various organizations, people are mostly involved in leisure interests groups (about 26 per cents) when compared to 4-13 per cent that are involved in religious,

NGOs and political organizations. Despite the fact that a majority of respondents confirmed that Lithuanians are too passive in the field of civil engagement, experience shows, that these issues are not personal priorities for them, individually.

Bearing in mind such low engagement numbers, it is obvious that society needs encouragement for more active participation. One of the ways to accomplish this is to present more opportunities for engagement in the resolution of civil issues using social technologies. An insufficient usage of social technologies is often identified as one of the main factors, why persons nowadays do not participate in socially oriented activities (Hampton, 2012; Foth et al 2011 and etc.). Social technologies no doubt create easier access for community members to participate in different decision making processes. This issue must be reasonably analysed since the fact, that social technologies are accessible to a majority of the residents. In the third quarter of 2013 more than 75 per cent of Lithuanian residents (age 16-74) were using computers and the Internet, with more than 28 per cent of residents using the Internet from their portable devices such as mobile phones or tablets (Information Society Development Committee under the Ministry of Transport and Communications, 2013). Moreover, Lithuania according to FTTH Council Europe, in 2012 still remains the leading country in Europe with the number of households using fiber-optic Internet. Taking into consideration the factors of low civil engagement and high Internet accessibility, it may be concluded that all opportunities for transferring civil engagement activities do exists. Thus despite this, it is worth to draw attention towards the existing situation in the field of virtual communities and networks in Lithuania. The growth of the number of virtual socially oriented communities is often observed. Therefore most new players that are not connected with government sector or other institutions authorised to make decisions, contributes to the fact, that a larger part of initiatives remain unrealized (Mačiulienė, Leichteris, Mačiulis, 2013). Of course such lack of functionality, leads to declination of trust. The mentioned issues thus provides a basis for exploring the field of virtual civil engagement more, with a focus on the reasons, that discourage individuals to involve themselves in virtual communities for social-oriented activities.

2.3 Opportunities of social technologies application for civil engagement

The above concluded propositions, actually present opportunities but not a guaranteed realization of broad application of social technologies for socially-oriented activities. In Lithuania usage of social technologies has one paradox: residents enjoy using technologies for work, leisure and personal every day needs, but most of them are not active users of various socially oriented platforms. If more than 75 per cents of residents (age 16-74) have access to Internet (IVPK, 2013), why are there 4-12 per cents of population involved in socially oriented activities (Vilmorus, 2012)? The low level of civil engagement in Lithuania has encouraged authors to perform theoretical research, which aims to identify the main risks of developing collective intelligence in network society (Skaržauskienė et al. 2013). The theoretical research showed, that frequent involvement in virtual communities meet risks such as information disclosure, infringement of privacy, threats for personal data, threats for intellectual property, censorship (possibly initiated by administrators or State) and other type of violations of rights and obligations (for example hate crimes, committed in virtual space). Systematic analysis of these five major risks, states that all of them are within the field of investigation of cyber security. According to the results of previous theoretical research, 13 prepositions for public opinion survey are presented (see Table 1).

Table 1: Prepositions for public opinion survey on cyber security related issues

No.	Proposition	Cyber security issue
1.	Participating in virtual communities is fully safe, as participation in social – political activities in real life	All issues
2.	Members of virtual communities should not be afraid of violation of their rights	Other type of violation of rights and obligations
3.	Personal data security is fully ensured during the participation in virtual communities	Threats for personal data
4.	The rights of intellectual property are fully observed taking a part in virtual communities	Threats for intellectual property
5.	The freedom of speech is fully guaranteed acting in virtual communities	Other type of violation of rights and obligations
6.	State controls the content of Internet	Censorship
7.	There is much more intolerance and defamation in virtual space to compare with real life	Other type of violation of rights and obligations

No.	Proposition	Cyber security issue
8.	The possibility to steal your identity exists in virtual reality	Threats for personal data
9.	The activities of virtual communities should be regulated by law in detail	Other type of violation of rights and obligations
10.	The information, provided in web sites of virtual communities may be controlled by administrators of web site	Censorship
11.	The administrators of virtual communities should take the responsibility of the content, which is spread	Other type of violation of rights and obligations
12.	Virtual communities must have strict and trustful system of authorization of members	Information disclosure / Threats for personal data
13.	Strict responsibility of virtual community members for the violation of rights of other individuals must be envisaged	Other type of violation of rights and obligations

3. Methodology and results of the research

3.1 Methodology

In striving to identify the importance of different legal threats of using such social technologies as virtual communities to promote the inclusion of society into decision making, we conducted a quantitative survey. A public opinion and market research company executed the survey in November – December 2013. The main goal of quantitative research was to ascertain the amount and level of involvement and participation of residents in the process of generating collective intelligence. 1022 residents were polled (478 male and 544 female) age 15-74 of all districts (cities and rural areas) of Lithuania, which guarantees a statistically reliable sample (with the confidence level of 95 percent) of the Lithuanian population. Public opinion survey was carried out using the method of direct interview at respondents' houses, using computers. For the households and respondents, random stratified selection was used. Interviewed respondents represented the overall Lithuanian population by the major socio - demographic characteristics (using stratified random sampling, with special reference to the participation in the shaping of collective intelligence factor).

3.2 Results of the research

An original research instrument was designed and included these main sections: I – the level of interest in social technologies; II – the level of knowledge about the virtual community projects tackling societal problems; III – the content of the process of participation in virtual community projects; IV – the level of satisfaction in virtual communication. Results of the survey once again confirmed that active exploitation of Internet resources by the Lithuanian population as 44 percent of respondents use the Internet every day, other 20 percent – at least once a week. On the other hand, it was confirmed that low civil engagement of the majority of those who use Internet (from 59 to 67 percent) for business and private communication, information search, entertainment, financial operations and only 21 percent put some contents for the public (for instance write comments, participate in discussions, create articles and blogs). Those who did not participate in social issues oriented virtually or social networks, justify their passiveness by inadmissibility to this medium of communication (40 percent), lack of interest (36 percent), lack of time (22 percent). Though it was believed that it is a waste of time (9 percent) as is the same important reason for the possible violation of privacy (9 percent of all answers), and almost as important as low level of security (6 percent). The importance of this might be confirmed by the fact, that more than half of the respondents who gave these answers are male and have higher education degree, which means they really understand the essence of risks and their importance. Security is almost an equally important for both users (19 percent) and non users (15 percent) of social initiatives. This contrast might be seen from Figure 1.

An in-depth analysis on the issues related to cyber security was mentioned in previous chapter. To evaluate these issues, respondents were asked to measure on the Likert scale the importance of the mentioned 13 aspects of cyber security as well as averaging of their opinions received. Figure 2, which compares all 13 aspects, shows how, varied the agreements and disagreements with the given statements and the average range of the opinions of respondents. The analysis of the highest four evaluated and four lowest evaluated propositions demonstrates that people are typically worried about threats for personal data and for intellectual property as well as violation of rights and obligations. The significant responsibility of virtual communities' members and virtual communities' administrator (both opinion average 4.1) is highlighted, meaning that any type of actions that strengthens this will have to be taken. Not of less importance is the

instance that users might be scared that their identity might be stolen (opinion average 4.1). Finally censorship is an important, key equal essential for detailed regulation for the risk of intolerance and defamation (all opinion average 4.0).

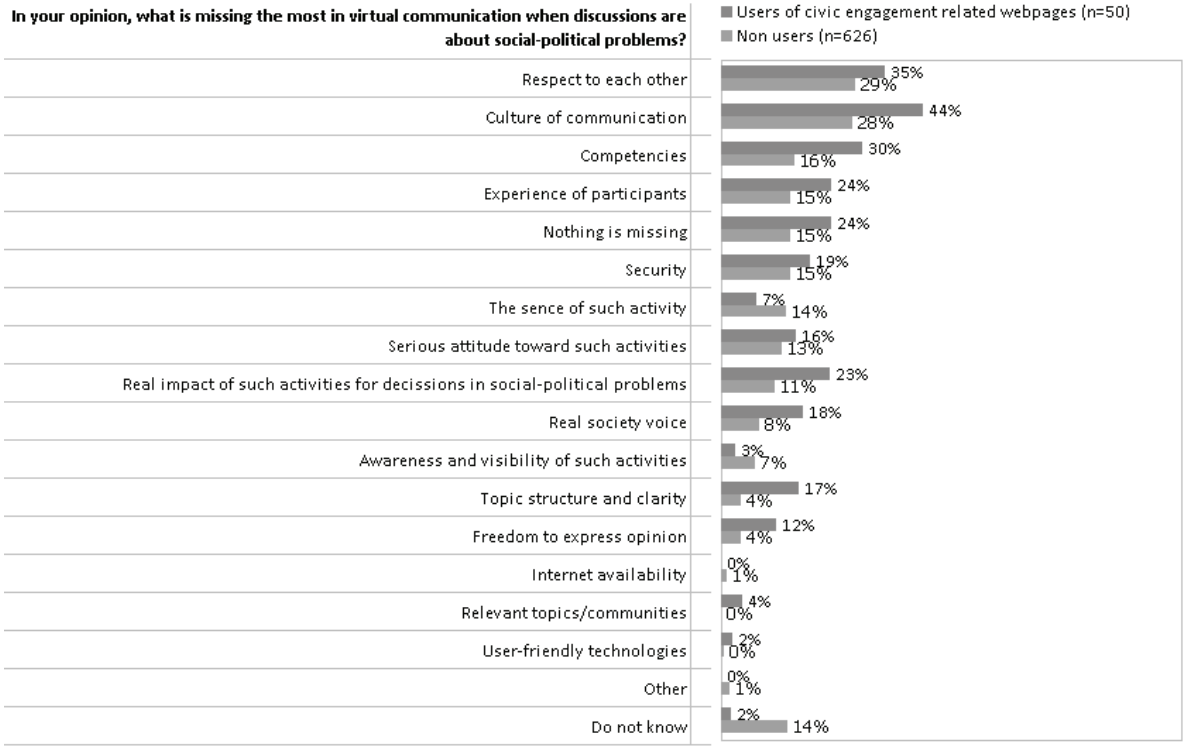
In your opinion, what is missing the most in virtual communication when discussions are about social-political problems?
Users of civic engagement related webpages (n=50)
Non users (n=626)
Respect to each other 35% 29%
Culture of communication 44% 28%
Competencies 30% 16%
Experience of participants 24% 15%
Nothing is missing 24% 15%
Security 19% 15%
The sence of such activity 7% 14%
Serious attitude toward such activities 16% 13%
Real impact of such activities for decissions in social-political problems 23% 11%
Real society voice 18% 8%
Awareness and visibility of such activities 3% 7%
Topic structure and clarity 17% 4%
Freedom to express opinion 12% 4%
Internet availability 0% 1%
Relevant topics/communities 4% 0%
User-friendly technologies 2% 0%
Other 0% 1%
Do not know 2% 14%

Figure 1: Reasons affecting low participation in virtual communication with purpose to discuss the social - political problems

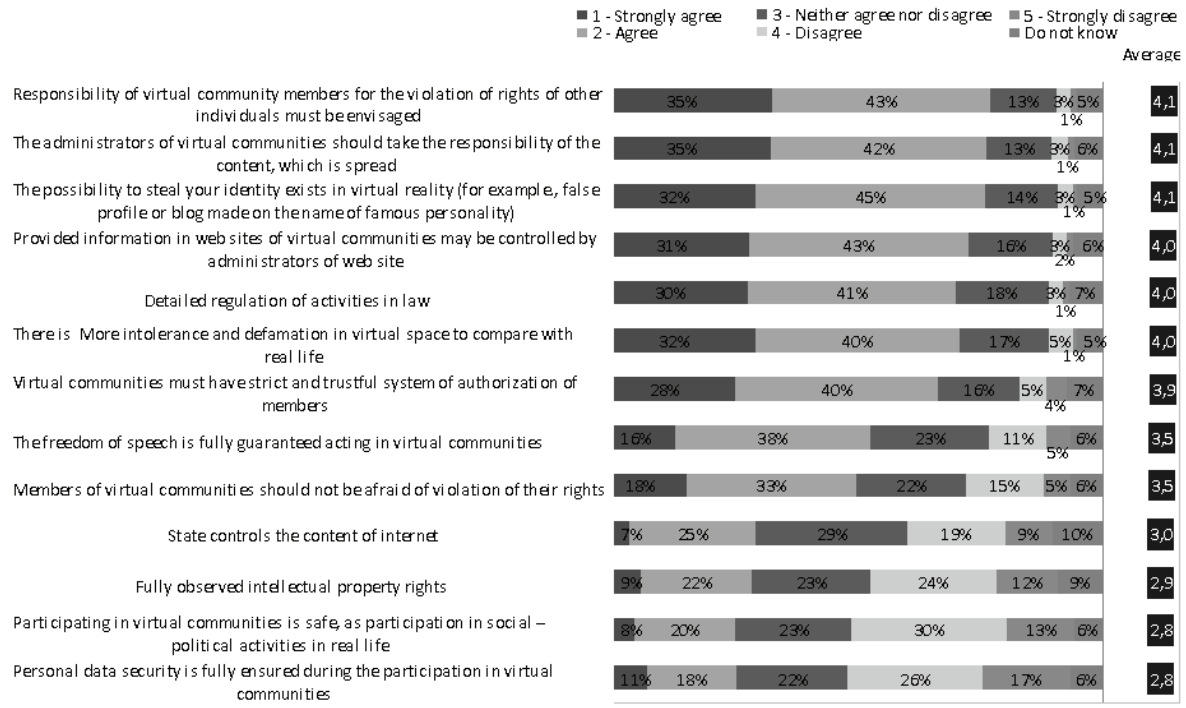
1 - Strongly agree
2 - Agree
3 - Neither agree nor disagree
4 - Disagree
5 - Strongly disagree
Do not know
Average
Responsibility of virtual community members for the violation of rights of other individuals must be envisaged 35% 43% 13% 3% 5% 1% 4,1
The administrators of virtual communities should take the responsibility of the content, which is spread 35% 42% 13% 3% 6% 1% 4,1
The possibility to steal your identity exists in virtual reality (for example, false profile or blog made on the name of famous personality) 32% 45% 14% 3% 5% 1% 4,1
Provided information in web sites of virtual communities may be controlled by administrators of web site 31% 43% 16% 3% 6% 2% 4,0
Detailed regulation of activities in law 30% 41% 18% 3% 7% 1% 4,0
There is More intolerance and defamation in virtual space to compare with real life 32% 40% 17% 5% 5% 1% 4,0
Virtual communities must have strict and trustful system of authorization of members 28% 40% 16% 5% 7% 4% 3,9
The freedom of speech is fully guaranteed acting in virtual communities 16% 38% 23% 11% 6% 5% 3,5
Members of virtual communities should not be afraid of violation of their rights 18% 33% 22% 15% 5% 6% 3,5
State controls the content of internet 7% 25% 29% 19% 9% 10% 3,0
Fully observed intellectual property rights 9% 22% 23% 24% 12% 9% 2,9
Participating in virtual communities is safe, as participation in social – political activities in real life 8% 20% 23% 30% 13% 6% 2,8
Personal data security is fully ensured during the participation in virtual communities 11% 18% 22% 26% 17% 6% 2,8

Figure 2: Results from public opinion survey on importance of cyber security related issues

Analysis of the four least evaluated possibilities also draws an interesting picture. The statements that respondents strongly disagree or disagree only shows the importance for State control over Internet content,

the lack of attention toward intellectual property in virtual communities and the fact that neither people feel safe in virtual space nor they think their data is secured. While evaluating the first statement they agreed with importance of these aspects. Where they disagreed was where their opinion was low on development of these aspects in Lithuania.

4. Conclusions

The growth of application of innovative technologies builds new possibilities and better quality social decision-making. Application of social technologies for collective intelligence that emerge in the activities of virtual communities is a new feature of civil engagement, which grants more effectiveness and compliance with society needs.

Even though good intentions build higher prospects, the realization and active use of socially oriented production in virtual space have certain risks. In previous theoretical analysis (Skaržauskienė et al. 2013), among the identified risks observed were cyber security related ones such as false identity issue, intellectual property issues, and censorship. The main legal problem identified was establishing a balance between privacy and the requirements for identification, as well as the positive outcomes between copyright law and the effect of synergy that are used in Collective Intelligence, between the need to control the content of communication in virtual medium in order to avoid the violation of human rights and the freedom of expression, which is the main advantages of virtual life (Skaržauskienė et al. 2013).

Following this analysis, the article focused on the analysis of legal concerns highlighted by scientists that restricts the broad involvement of virtual communities in civil activities. Theoretical analysis assisted in the concluding of propositions and presented opportunities but not a definitive understanding on the broad usage of social technologies for socially oriented activities.

The public opinion survey done for the research , paid close attention to the key aspects of active users of virtual communication or reasons that keep away non-users from using virtual tools for civil engagement. The national context analysis might be useful for answering why initiatives and models for protection from cyber-attacks are becoming less effective despite the efforts made, as well as emphasizing the reasons of low civil engagement via virtual tools. The results demonstrate that respondents are mostly worried about threats for personal data and for intellectual property as well as about violation of rights and obligations. Yet, respondents stated the importance of State control over content of the Internet, the lack of attention toward intellectual property in virtual communities as well as the fact that neither people feel safe in virtual space nor they think their data is secured. So the research provided a better understanding of different aspects that caused paradoxical usage of social technologies for work, leisure and personal daily needs but not for civil engagement.

These answers clearly emphasize that attention must be paid towards cyber security issues in order to develop safe and reliable environment for people, who wish to engage and wish to generate ideas for greater welfare of society. The research results also prove the necessity for a deeper qualitative analysis to identify the actions, possibilities and motives of the other stakeholders (e.g. virtual communities' creators, administrators, active users etc.). These are the essential knowledge needed to minimize cyber security related risks required for fostering engagement of society and monitoring the CI emergence in networked platforms that enables significant changes in civil engagement.

Acknowledgements

The research is funded by European Social Fund under the initiative „Support to Research Activities of Scientists and Other Researcher (Global Grant)", and administrated by Lithuanian Research Council (grant No. VP1-3.1-ŠMM-07-K-03-030, project title "Influence of Social Technologies on the Development of Collective Intelligence in Networked Society'').

References

Adler, P. R., Goggin, J. (2013) "What Do We Mean By "Civil Engagement"?" *Journal of Transformative Education*, Vol. 3 No. 3, July, pp. 236-253.

Castells, M., Cardoso, G. (eds.) (2005). *The Network Society: From Knowledge to Policy*. Washington, DC: Johns Hopkins Center for Transatlantic Relations

Chinn, D., Kaplan, J., Weinberg, A. (2014) Risk and responsibility in a hyperconnected world: Implications for enterprises [online]. Report from the World Economic Forum and McKinsey & Company http://www.mckinsey.com/Insights/Business_Technology/Risk_and_responsibility_in_a_hyperconnected_world_Implications_for_enterprises?cid=other-eml-alt-mip-mck-oth-1401

Foth, M. (2011) *From Social Butterfly to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagement,* MIT Press, Cambridge.

Grady, M, Parisi, F (2006) *The Law and Economics of Cybersecurity,* Cambridge University press, Cambridge.

Gunawardena, C. N., Hermans, M.B., Sanchez, D., Richmond, C., Bohley, M., Tuttle, R. (2009). A theoretical framework for building online communities of practice with social networking tools. *Educational Media International*, Vol. 46, No. 1, p. 3-16.

Hampton, K.N. (2011) "Comparing Bonding and Bridging Ties for Democratic Engagement", *Information, Communication & Society*, Vol. 14, No. 4, June, pp. 510–528.

Jokubaitis, M. (2012). Lithuania Again Gains Name of Internet Speediest [online]. FTTH Council. http://www.ftthcouncil.eu/documents/Clippings/2012/February/20120217-Lithuania.pdf

Mačiulienė, M. , Leichteris E. and Mačiulis A. (2013) "The perspectives of developments of virtual communities in Lithuania" [online]. *Social Technologies'2013 conference proceedings*, http://www.mruni.eu/mru_lt_dokumentai/fakultetai/socialiniu_technologiju_fakultetas/SOCTECH_2013_proceedings.pdf

Malone, T.W, Laubacher R. and Dellarocas C. (2010) "The Collective Intelligence Genome" *MIT Sloan Management Review,* Vol. 51, No. 3.

Morozov, E. (2013). *To Save Everything Click Here,* Public Affairs, New York.

Public Opinion Survey "The Concept of Public Spirit" [online]. Vilmorus. http://www.3sektorius.lt/docs/Ataskaita_pilietiskumas_santrauka_2013-01-17_15_59_37.pdf

Salminen, J. (2012). Collective Intelligence in Humans: A Literature Review [online], *MIT, Collective Intelligence 2012*, http://arxiv.org/abs/1204.3401

Shoemaker, D, Conklin, Wm. (2011) *Cybersecurity: The Essential Body Of Knowledge,* Cengage Learning, Farmington Hills.

Skaržauskienė et al. (2013) "Risk and Opportunities for Developing Collective Intelligence in Networked society" [online]. *Social Technologies'2013 conference proceedings*, http://www.mruni.eu/mru_lt_dokumentai/fakultetai/socialiniu_technologiju_fakultetas/SOCTECH_2013_proceedings.pdf

The Review of Development of Information Society (2013) [online]. Information Society Development Committee under the Ministry of Transport and Communications. http://www.ivpk.lt/uploads/Statistika/2013%20m.%20III%20apzvalga.pdf

Tvaronavičienė A. and Paražinskaitė G. (2013) "Legal Presumptions of Usage of Social Technologies for Virtual Communities's Involvement in the Decision Making Process in Lithuanian Republic Legal System" [online]. *Social Technologies'2013 conference proceedings,* http://www.mruni.eu/mru_lt_dokumentai/fakultetai/socialiniu_technologiju_fakultetas/SOCTECH_2013_proceedings.pdf

Determination of Meme Proliferation Factors

Namosha Veerasamy and William Aubrey Labuschagne
Council for Scientific and Industrial Research, Pretoria, South Africa
nveerasamy@csir.co.za
wlabuschagne@csir.co.za

Abstract: The ubiquity of the Internet has become part of everyday life. The Intenet allows for the searching of information on topics that are relevant at that particular time. The possession of valuable information by members of a network improves the worth of such members and subsequently increases their influence. Memes refer to pieces of information or message that replicate. Social networking sites on the Internet provide an ideal platform to propagate messages as groups of like-minded people connect with one another and communicate ideas. Messages pertaining to popular, trending or controversial topics become memes on social networking platforms. The design of memes serving as narratives has the potential to create disinformation or create a revolution as in the case of the Arab Spring. Narratives are the use of stories to influence behaviour. Messages could be created for nefarious purposes and if these messages become memes, these malicious messages can replicate at a rapid rate and potentially influence behaviour and action in a negative manner. The converse is also true a positive change could be attained though the propagation of memes designed for that specific outcome. It is essential to understand what factors are required to effectively disseminate memes within a narrative network with the goal to transform behaviour of the members in the social network. As the majority of Internet users indiscriminately trust the content encapsulated within the information highway, a threat exists in which entities with power could influence society. This paper addresses factors that could increase the proliferation of memes within cyber space. A framework is proposed of high-level factors that can be used in the design of a meme so as to increase its proliferation. Thereafter, the paper presents an experiment whereby these proliferation factors were investigated. Overall, the goal of the paper is to demonstrate how messages can be designed to proliferate and potentially influence individuals.

Keywords: influence, narratives, meme, profile, propagation, social networks

1. Introduction

Narrative networks form through the use of stories that helps create identity and belonging in a community with shared interests. Already in 2001, Arquilla and Rondfeldt (2001) explained that some narratives can be used for manipulative purposes by trying to create connection in a network. Arquilla states that disinformation, misrepresentation and outright lying are downsides that can result from the narrative level. The creation of unscrupulous storylines can unfold in the media. The 9/11 terrorist attacks are a typical example of a potentially negative narrative. Some claim that it was not a terrorist attack but a cover-up from the US government (Woodworth 2014). Similarly, great speculation rages over the cause of the Malaysian passenger plane that disappeared from skies at the beginning of 2014 (Quinn, Branigan 2014). Some part of society could use narratives to frame this event as a conspiracy by the Malaysian government (Adams 2014).

In 2011, the Defense Advanced Research Projects Agency (DARPA) welcomed proposals for the creation of narratives (books, propaganda, stories, that cause people to think) in order to create concealed messages that would influence vulnerable people (Yirka 2011). DARPA is the Pentagon's division responsible for the development of cutting-edge technology for the USA Department of Defense (Heyes 2012). The proposal requested narratives that could have a powerful influence on human thoughts and behaviour by influencing memory, emotions, judgment bias and personal identity (DARPA 2011). Narratives could potentially influence political radicalisation, violence in social movements and the course of insurgencies. Post (2007) mentions that the Internet has helped to propagate the ideology of right-wing extremism, as well as Islam. Individuals can become enraged by extremist messages and thereafter even resort to computer based attacks. Essentially narratives stemming from the Internet and social networks have the potential to serve the role of weaponisation. Narratives could serve as a weapon to hijack the mind and plant false but believable stories (Heyes 2012).

The evolution of the Internet has given viral messages an enormous boost (Chielens 2002-2003). Movie and television stars, sports, scandals, scientific discoveries and incredible animals are examples of popular topics that go viral over the Internet. However, messages could also be created for malicious purposes and if these messages become memes, these malicious messages can replicate at a rapid rate and potentially influence behaviour and action in a negative manner. A meme is defined as "*an idea, behavior, style, or usage that spreads from person to person within a culture*" (Merriam-Webster 2014). Melodies, catch-phrases, fashion,

and the technology of building arches are seen as examples of memes (Dawkins 2006)**.** In this research context, the term meme and message is used interchangeably, to convey ideas that could influence another person. In other words memes have now become a popular form of cultural transmission and its main premise is to be memorable, which helps promote its replication.

Blackmore (2000) has explained that people spread memes indiscriminately irrespective of whether they are useful, neutral or harmful. She states that a marvellous scientific idea or technological invention may spread because it is useful but a song like Jingle Bells is not useful but spreads because of its melody. Other memes could also be harmful like chain letters, pyramid schemes or promotions for slimming diets and dangerous medical cures. Thus, by gaining knowledge of what makes a meme, one can also help create awareness against dangerous memes.

Cialdini (1998) discusses the premise of acceptable norms. He explains that preferences are established within the social network and any deviant behaviour will be discouraged by members (Cialdini, Trost 1998). Consequently, if negative norms are created within a social network, malicious behaviour can be promoted and condoned. As is the case with terrorist organisations, negative narratives on their websites and social networks can encourage violent behaviour and dissident acts of protestation and violence. Similarly, the media or other interest groups can create narratives on social networks in the form of memes in order influence behaviour and opinion.

Social media makes it easier to tell stories that contribute to narrative psychology (Renando 2011). Through social media, individuals share their recollections and get feedback, which essentially contributes to narrative psychology. Through these interactions, social media helps develop meaning and identity. The events of the Arab Spring uprising depict the influence of narratives in the social media. Social networking sites were used by activists to publicise the protests in Egypt. Social media- its rise and new activist uses have played a critical role in mobilisation, empowerment, shaping opinions and influencing change (Huang 2011). Therefore, the creation of memes through narratives can help influence behaviour, opinion or viewpoints.

To understand the dynamics of meme propagation on social networks, it is important to study the dynamics of adoption. Kempe, Kleinber and Tardos (2003) state the premise of viral marketing is initially target a few influential members of the network. Similarly, in order to create a meme it will be important to find influential members in a network. Other dynamics can also create adoption and thus the focus of this study will be to determine other factors that influence meme proliferation.

In social networks, users build their social currency and influence by increasing their number of friends/followers together with frequent updates about activities, opinions and interests. If a message is sent out and a personalised response is returned, trust can be strengthened. Automated responses may be unhelpful or suspicious. Thus, the nature of responses and updates can create trust. When an active or popular user has many friends and frequently updates current or useful information, these messages can begin to proliferate at a rapid rate. Memes are considered as information that replicates (Aunger 2002). Popular or trending topics can replicate and then become memes as they are created from the most prevalent or entertaining issues.

Often for a meme or trend to begin, the message needs to be initially propagated to other members of the network. An inactive account or account with no real friends will hardly create enough trust. In addition, the nature of the message also needs to be considered. However, various factors can influence the proliferation of memes. Therefore it is imperative that in order for messages to become memes or trends, certain conditions are present.

This paper discusses the findings from an experiment conducted to determine what initial factors could be beneficial to the development of a narrative network which would utilize memes to effectively disseminate information to members of a network. Initially a framework is proposed that discusses factors that contribute to the proliferation of a meme. Thereafter, the design and results from the experiment are given.

2. Proposed high level framework for meme design

In this section, we address the high-level design of the memes for the proposed experiment. If a person mentions or retweets a message this shows awareness or interest in the issue. Therefore, a mention or retweet can be used to create a meme and stimulate interest and awareness in an issue. A retweet or mention can help attract more followers or proliferate the meme.

To design a meme to be attractive the following factors should also be considered:

- **Newness** – New content or breaking news has a higher probability to propagate between users in a network as users who possess new information have higher status within social groups.
- **Emotiona**l – Memes that evoke positive emotions, for example happiness and hope could be beneficial to other users. Conversely, the opposite is also true as negative emotions such as in the case of a natural disaster or breaking news could be used to propagate through the network and subsequently inform and influence other users.
- **Enticing** – The use of wording to draw the attention of the reader to the message. The words include "WOW","OMG" and "Shocking".
- **Targeted** – Social networks are created around people who are like-minded and have the same interests. A users within a specific social network build around a specific interest is higher in value if information is disseminated which benefits the group.
- **Richness** – A message can be a combination of text, links to external resources on the Internet and images. The presence of all these would be more of value to the user as more senses of the user are used, which enhances the experience.

Veerasamy and Labuschagne (2013b) explain to determine trust in a social network account, one should consider the activity, friends, information, albums and profile picture of the account. Essentially by studying the profile and activity, one gains information about the behaviour and personality behind the account. This helps build trust in an account. Interaction with an account creates trust and a person is more likely to resend, like or favourite a message from a known account. When a person has many contacts, their social status increases and this may create influence. Influence and enticement can be created through emotive or positive words.

Advertising principles can also provide great insight into the marketing of a topic and the creation of a meme. In his book Contagious Why Things Catch On, Berger (2013) describes a model entitled STEPPS which stands for Social Currency, Triggers, Emotion, Public, Practical Value and Stories. A brief explanation of these insightful principles follows.

The first part of the STEPPS model refers to Social Currency. Social Currency refers to increasing a person's social status by being knowledgeable about trendy and interesting topics. Social currency can be created through bringing out the novelty through stories and exaggeration. The use of games and loyalty schemes also creates interest. Social currency is created when consumers boast about scheme benefits and promote the game play.

The second principle of the STEPPS model is Triggers. Stimuli and triggers from the environment can help remind people of a certain concept. Frequent stimuli, word association, alliteration and idea linking are all techniques that can trigger a response. In some cases a negative response like a critical review can generate interest as it creates attention to the topic.

The third principle that can be applied from the STEPPS model is the use of Emotion. Emotive words can appeal to the gentle side of people. On the other hand negative words also arouse emotion like anger or outrage. Remarkable topics like scientific discoveries can also create the emotions like awe and amazement. Humour is also a very popular emotionally-enticing technique.

The next principle in the STEPS model is a very simple one: Public. If a concept is not in the public eye it cannot be seen. A rather obvious deduction, it is important to note that if a topic is not placed on a platform that can be seen by many, its message can be lost.

The second P in the STEPPS model stands for Practical value. Tips for saving time, improving health, or saving money will spread as it provides consumers with a beneficial service. Incredible offers, special deals, health, reviews and education articles are all useful and are thus popular topics to spread.

The last principle from the STEPPS model is Stories. Many people are epic story tellers. The interest that is created out of stories can be used as a marketing advantage. The principle of STORIES can incorporate the other principles of the STEPPS model as well. Stories can be told to create social currency (by trying to look impressive or create envy), for practical value, with emotional arousing tones, as recounts of awe-inspiring tales or to create lessons. Overall, to use this technique make the product/topic part of the story in order to create interest.

Now that the STEPPS model has been explained, the discussion moves on to other design principles that can be applied. Blackmore has explained that memes form from variations and combinations of old ones, either in a person's mind or when memes are passed between people (2000). Furthermore, Young (2012) has stated that coming up with a completely new idea to go viral may be difficult and in some cases just a matter of luck. Young encourages the use of borrowing an idea that is already popular.

Quodosi (2012) talks about different techniques of creating memes that generates interest. These include:

- **Communication**- The methodology of memes relies on a simple message that is easy to grasp. Memes are not very text heavy
- **Image**- Most memes are sent using an image. However, Bernstein (2012) has also shown that videos are very powerful and can also go viral. He explains that many companies succeed in creating memes by using YouTube in their social media advertising strategy.
- **Appeal**- memes provides some type of appeal which includes fascination to our ego, or amusement and distraction.

Garun (2012) explains that following the right people is a great filter for finding content as it helps see what is viral. By using the judgement of friends and trusted sources, the credibility of the content is increased and promotes viewing the content. Therefore, it is imperative to realise that people rely on the judgement of their trusted contacts.

Moreover, Brandon (2013) discusses various ideas that are helpful for the high-level design of memes that will be used during the experiment (See Section 2):

- **Post the same links multiple times**- Many users only carry out "drive-bys" on social networks streams and can thus miss viewing some content. It is better to post a link a few times at different zones to ensure that more potential followers can be reached.
- **Reciprocate followers and also follow them on social channels**- Some followers may not be aware of a person's handle on a certain channel so it helps to be visible in many channels. Try to users with similar interests or those that are "influencers" (have many followers). A good strategy is to follow influencers as it can help find another league of new followers as well.
- **High frequency of interaction with followers**- When regularly interacting with followers a relationship can be established. When big corporations provide responses to message, assurance in the brand is built. Similarly, recurring messages helps establish trust and create an open channel for communication.
- **Show personality**- Automated or random messages can seem impersonal or suspicious. It may seem like a bot has generated them. When creating contacts it is important to show that the user is legitimate and will be using the media frequently. Answer messages directly, follow-up responses or using humour can all establish users as an "influencer". Staying active and continuous engagement is critical.

Furthermore, Chielens (2002-2003) has stated that in an ideal situation, a replication will have exponentially increasing number of carriers which serve as hosts/vectors for this replication. Public awareness is essential for creating a viral concept. Godin (2013) explains that a critical aspect of creating a virus is the source of the viral infection. For a virus to proliferate, it needs to come into contact with people. Similarly for a meme to go viral, it needs to have a large number of people to reach. In addition, Godin explains that if there is a specific idea to spread, find the matches that would also be interested in the idea. Targeted message may create a bigger impact as they reach the appropriate audience.

Based on the discussion in this section, Figure1 shows a high-level summary of the design factors for the proliferation of a meme. Due to the time restrictions of the experiment only three of the STEPPS principles are utilised in the framework. (Social Currency, Practical Value and Story-telling). This framework refers to several terms used in the Twitter environment. However, a subset of the principles encompassed can be applied to other social networks. For example, a Retweet correlates to re-posting a message. Followers refer to the number of contacts. "Favourited" is term found on various social networking sites.

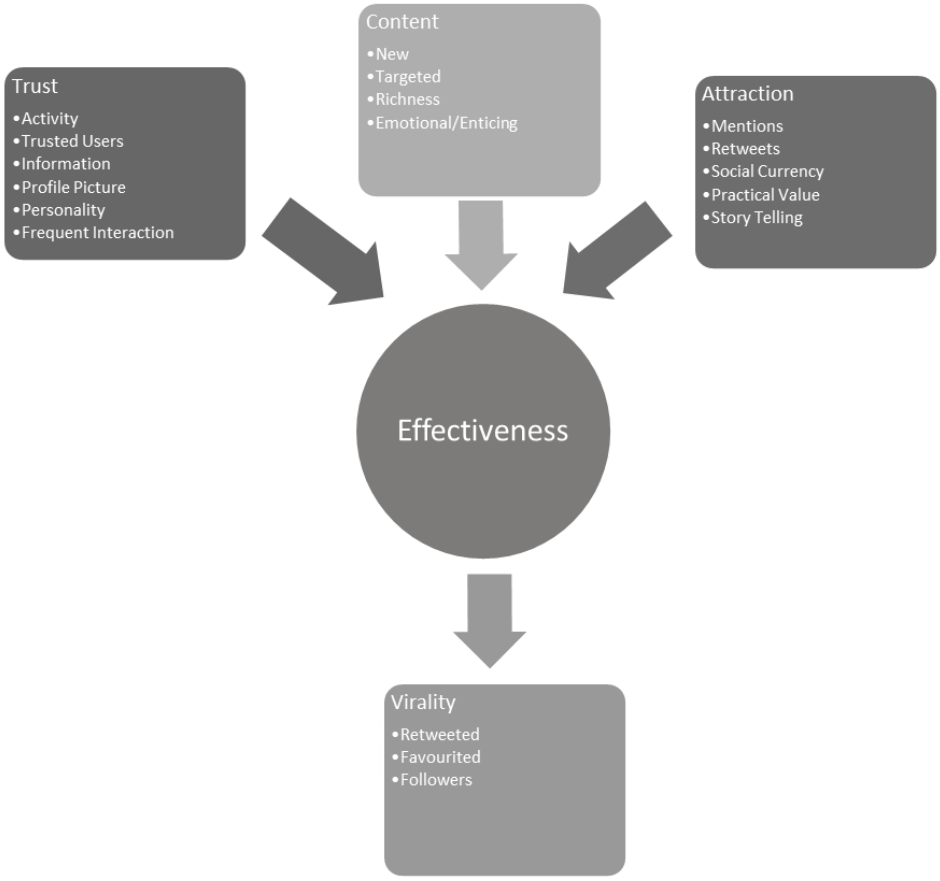

Figure1: Meme design framework

This section discussed various factors that can be applied when designing a meme. The next section will therefore discuss the application of the framework in an experiment.

3. Experiment

An experiment was designed to investigate the proposed design factors for meme proliferation. The next few sections describe the design, data collection and analysis of the experiment.

3.1 Design

An "A/B" test was used during the experiment. The main objective of using the "A/B" test is to determine if certain factors were affecting the outcome of each test. The experiment required to be conducted within an established environment with maximum exposure to the Internet society and provide a mechanism to collect data generated during the experiment. The social media site Twitter was selected to conduct the experiment as it provides an effective platform to disseminate memes to all the users. Twitter is used as a broadcasting platform where all Twitter users can view public tweets. A tweet in the context of Twitter is a message that contains 140 characters to convey information in the format of text and in some instances images. In addition Twitter also provides researchers with access to its tweets via an Application Program Interface (API).

Two Twitter accounts "supr3m3travel" and "jillhilljackson" were created for the experiment. These accounts would be individually used to tweet memes that were designed to convey different ideas. Also, the accounts

were personalized as this increases the creditability of the account (Veerasamy, Labuschagne 2013a)**.** For example "supr3m3travel" was themed around the topic of travelling and "jillhilljackson" was generalized. In other words, the building of trust within social networks is imperative to the success of gaining a following on social networking sites, which in turn would improve the chances of a meme becoming viral. A newly created profile on Twitter does not have a profile description or a profile image. People are less likely to trust this profile and not follow the account until actions are performed on the profile to improve the trust levels.

Next a developer account was created for each newly created Twitter account to access the Twitter API (Twitter 2014). This resulted in the capability to extract data about each account as a single account does not allow access to data created by other Twitter accounts. Hence, "supr3m3travel" cannot access data about a tweet created by "jillhilljackson". For example, one of the methods, "*GET statuses/retweeters/ids"*, lists the users who retweeted (RT) a tweet created by a specified user (Twitter 2013). However, if "jillhilljackson" created the tweet then "supr3m3travel" cannot list the users who retweeted the tweet created by "jillhilljackson".

A data collection system was programmatically developed, which would be executed weekly to connect and extract tweet data from the Twitter data sources via the API. In the context of this research, only the creation date, the identification number (ID) of the tweet, the number of retweets, the number of favourites, as well as the number of followers per tweet were collected. The unique ID of the tweet is used as an input to determine the users who have retweeted the tweet. Hence, this would only be determined if the number of retweets for the tweet was higher than zero. A user who retweets a tweet finds the information useful to spread to their network. The number of favourites is an indication of users who found this tweet valuable for future reference and the number of followers indicates the size of the users following. These three metrics provides insight into the influence of the user and subsequently how effective the meme was disseminated. The collection of data was executed weekly as this would allow time for the meme to spread through the social network as Twitter users are dispersed around the globe and operate in different time zones.

A schedule together with the strategy was created for the experiment to conduct an A/B test (See Table **1**). This provided a high level description of which actions would be taken during the specified time intervals.

Table 1: Experiment A/B testing

Account	A		B	
	Strategy	Dates	Strategy	Dates
supr3m3travel	No interaction with other profiles Message specific to travel	2013/10/29 to 2013/11/20	Interaction (Reply reciprocation and Follow) Message specific to travel	2013/10/29 to 2013/11/20
jillhilljackson	Interactions (Reply) Message covering all fields	2013/11/21 to 2014/01/07	No interaction with other profiles Message covering all fields	2013/11/21 to 2014/01/07

Memes which focused on travel were created for the Twitter account "supr3m3travel" during the first phase of the experiment. Finally, the memes to be tweeted by each Twitter account were created using some of the factors according to the specified strategy in Table **1**. It is challenging to encompass the various design principles when facing word restrictions in Twitter. Applying story-telling, practical value and social currency principles, together with mentions, requests for tweets, emotive words, in conjunction with target specific requirements can be rather difficult using a limited number of characters. In a traditional advertising campaign, word limits may not be so stringent and thus the various principles of the STEPPS model and the design framework can be applied more advantageously. Thus, during this experiment the design factors were applied as far as possible in the context of the environmental restrictions.

The next section describes how the data was collected during the experiment.

3.2 Data collection

The first test (Test A) for both Twitter accounts was initiated on 29 October 2013 and completed on 20 November 2013. The strategy for the Twitter account "supr3m3travel" was to create meme around the theme of travel. This was aligned with the description of the account and subsequently would have increased the account's creditability as experienced by other Twitter users. Also "supr3m3travel" did not interact with other

Twitter account. In other words "supr3m3travel" did not retweet tweets created by other Twitter users or replied on other tweets. The other Twitter account "jillhilljackson" in turn interacted with other tweets created by other Twitter users and created general messages that were not specific to a topic.

Test B started on 21 November 2013 and concluded on 7 January 2014. The strategy was changed for both Twitter accounts during this phase. During this test the account "supr3m3travel" not only replied to other users but also followed them. The account "jillhilljackson" posted memes that were general in nature. Also no interactions were pursued by the account "jillhilljackson".

The data collection system described earlier was used weekly to connect to Twitter through the API and extract the requested data. This in turn was sanitized as only a subset of the data was required for this experiment and finally stored within a storage platform. The next section discusses the analysis of the results.

3.3 Analysis

The data sets for both Twitter accounts used during the experiment were graphically illustrated in Figure **2**. The data depicted in the figure shows the temporal data collected from 29 October 2013 to 7 January 2014. This shows the number of followers for each Twitter account at daily intervals. Figure **2** shows the growth in the number of followers on the "supr3metravel" account during the B test. During the B testing, the strategy for the "supr3metravel" account was to interact more frequently and follow new contacts.

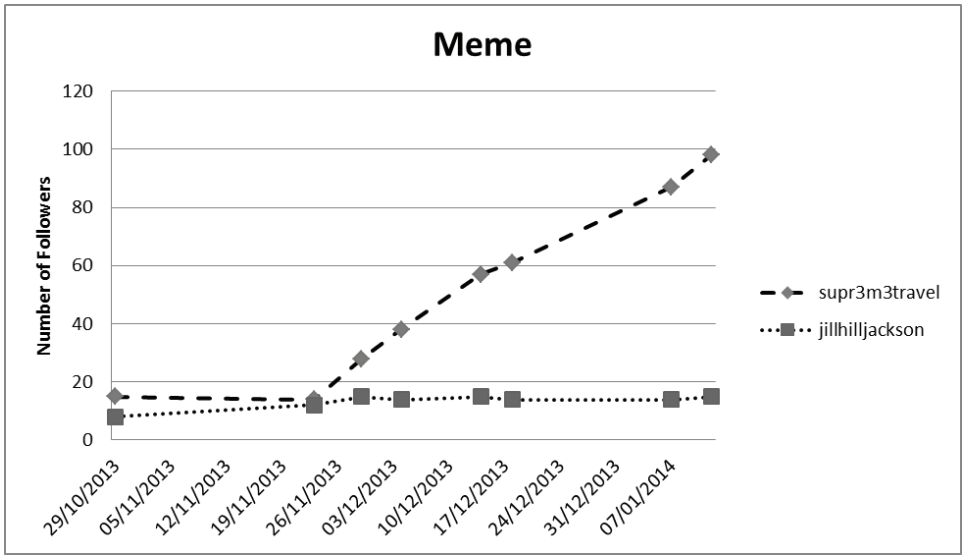

Figure 2: Growth of number of followers

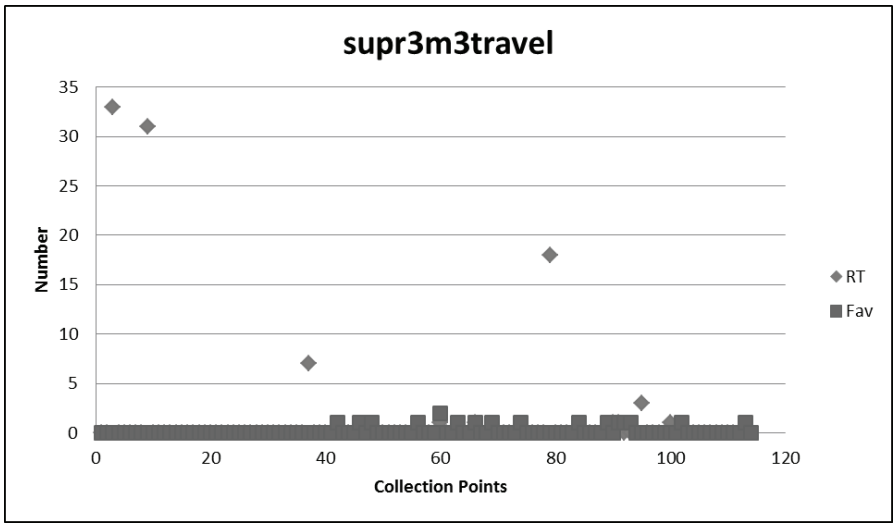

Figure 3: Number of Retweets and favourites on "supr3m3travel account"

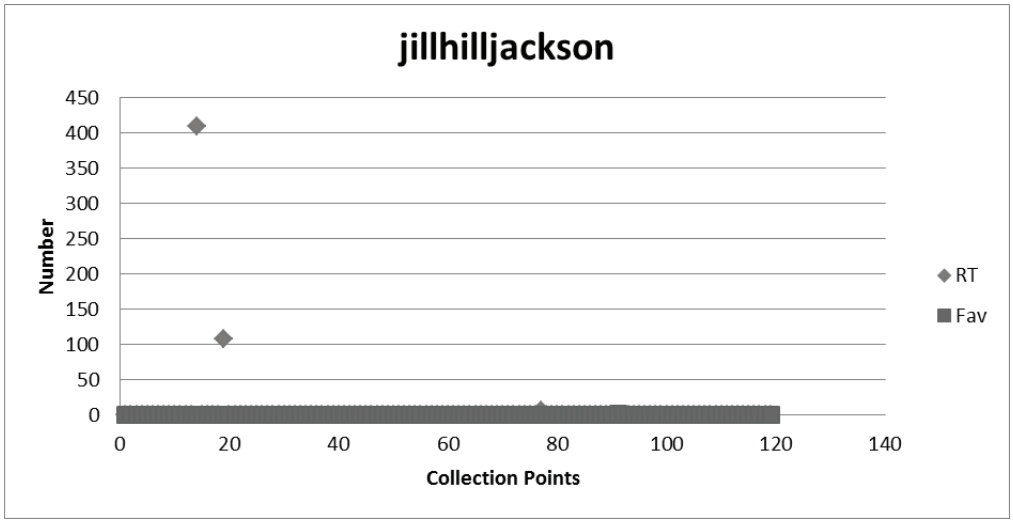

Figure 4: Number of Retweets and favourites on "jillhilljackson" account

A low number of retweets and favourites for each tweet created by the two Twitter accounts, were observed from the analysis of the results. However, the number of followers increased substantially for the account "supr3m3travel". This occurred during the second phase of the experiment. The strategy followed by the Twitter account "supr3m3travel" was to follow other Twitter users, together with the interaction on tweets posted by other users. This strategy showed a significant impact. The strategy resulted in a substantial increase in the number of followers during the second phase (Test B) of the experiment.

Travel related messages were posted on the "supr3metravel" account throughout the experiment. Similarly, generalised messages were posted on the "jillhiiljackson" account. Thus, the only metric that was different was the increased frequent interaction and following of other users. This finding provides valuable information to individual users of social networking sites like Twitter whose objective would be to create a following which subsequently could be used to exercise influence on. Another interesting finding was that although the number of favourites per tweet was low, the frequency of other Twitter users labelling the tweets as favourite also increased as the number of followers increased. This indicates that memes posted by the "supr3m3travel" account during that time had value to the users within its follower's network. In the experiment, not many contacts were initially made in the A test and thus this impeded the ability to have the message viewed and shared. However, during the B test, the number of contacts on the "sumpr3metravel" was increased by interacting and following new contacts. This is turn resulted in a reciprocal following of "supr3m3travel" and subsequently led to cluster of retweets and favourites.

Chielens (2002-2003) has explained that memes have to go through a "battle for the brain" and since there are a limited number of meme-carriers and the large amount of memes, there is a strong competition among memes. This results in some memes being passed on rapidly and others simply are fading away without any attention.

4. Future work

The focus in this paper was on the design and implementation of factors for meme proliferation. Future work will entail investigating the effect of specific factors to determine whether they create a significant influence. This will help in the design of targeted and effective memes. Furthermore, the study can be expanded to the topic of security awareness to determine whether a positive influence can be created using memes.

5. Conclusion

Integration between society and the Internet has achieved critical mass. A member of society relies on the presence of the Internet to conduct daily activities which include leisure and work. As most humans are gregarious in nature the quest for information to improve social status has increased. The Internet provides copious sources of information generated by society. With the birth of Web 2.0 social networks provided a platform which allows for content creation between like minded people on any related topic. It is also in the human nature to implicitly trust the information obtained from reliable sources. However, the Internet is faceless and subsequently not all information is equal and can be nefarious in nature. Narrative networks have

been proposed as a mechanism to influence society. The Internet together with the social networks is ideal platforms to implement narratives networks as part of a propaganda campaign and ultimate behavioural change. The use of memes could improve the successes of narrative networks.

The experiment investigated the initial factors required to establish trust and proliferate the developed memes. An unexpected finding highlighted that reciprocity especially on social networking sites could be used to establish trust. Subsequently the creditability of the user would increase to other members of the network. During the A/B test (the second phase), the account "supr3m3travel" saw a drastic increase in the number of followers once the user "supr3m3travel" interacted with other users and also actively followed other users. The effect of the increase in followers for the user "supr3m3travel" also demonstrated that content posted was perceived as valuable to the network as other users labelled the tweets as a favourite. In other words the influence of the user "supr3m3travel" increased in the network. Conversely the other account "jillhiiljackson" did not interact with other members in the network, the number of friends did not increase nor was any tweets deemed as valuable.

The success of memes not only lies within the design but also the platform where the memes are disseminated on. Social networking sites have exploded in the last few years and infiltrated society. Twitter provides a suitable platform for meme propagation as its design is developed around broadcasting to all users of the platform. The establishment of trust is fundamental to the success of memes. In other words members of a network would not share information from an untrusted source to other members. This helps keeps creditability intact. Trust factors are described as portraying human-like behaviour, for example having meaningful conversations, posting regular content, personalizing the user profile and having a legitimate following. The value of the user within a social networking is increased when relevant contributions are made to the network thus increasing the status of the individual and the group. The effectiveness of the meme is further increased if the content is original and has value in terms of usefulness and scarcity. In addition the evocation of emotions within the appropriate context would also increase the success of the meme.

References

Adams, M. 2014, *Malaysian government admits altering MH370 pilot transcript, hiding evidence and misleading the public in massive cover-up* [Homepage of Natural News], [Online]. Available: http://www.naturalnews.com/044542_MH370_Malaysian_government_concealing_evidence.html [2014, 0407].

Arquilla, J. & Ronfeldt, D.F. 2001, *Networks and Netwars: The future of terror, crime, and militancy,* Rand Corporation, Santa Monica, California, USA.

Aunger, R. 2002, *The Electric Meme,* New York: Free Press, New York, USA.

Berger, J. 2013, *Contagious: Why things catch on,* Simon and Schuster.

Bernstein, M. 2012, 28 May 2012-last update*, What Viral Videos Can Teach Us About Building A Brand* [Homepage of Memeburn.com], [Online].

Blackmore, S. 2000, *The meme machine,* Oxford University Press.

Brandon, J. 2013, 18 September 2013-last update*, 4 More Ways to Boost Your Twitter Followers* [Homepage of Inc.com], [Online]. Available: http://www.inc.com/john-brandon/how-to-hit-5000-followers-on-twitter.html?cid=sf01001 [2013, 21 December].

Chielens, K. 2002-2003, *The Viral Aspects of Language*, Universiteit Brussel.

Cialdini, R.B. & Trost, M.R. 1998, "Social influence: Social norms, conformity and compliance.", *Annual Review of Psycology,* vol. 55, pp. 591-621.

DARPA 2011, *Broad Agency Announcement Narrative Networks* , DARPA, United States of America.

Dawkins, R. 2006, *The selfish gene,* Oxford university press.

Garun, N. 2012, 15 May 2012-last update*, Manufacturing the Meme: Can you Create Web Virality?* [Homepage of Digital Trends], [Online]. Available: http://www.benchmarkemail.com/blogs/detail/small-business-strategy-understanding-memes [2013, 20 November].

Godin, S. 2013, 15 July 2013-last update*, Q&A: Controlling the IdeaVirus*. Available: http://sethgodin.typepad.com/seths_blog/2013/07/qa-controlling-the-ideavirus.html [2013, 20 November].

Heyes, J.D. 2012, *Pentagon researching 'narrative networks' as way to hijack the brain with false stories*, Natural News.

Huang, C. 2011, *Facebook and Twitter key to Arab Spring uprisings: report*, The National.

Kempe, D., Kleinberg, J. & Tardos, É. 2003, "Maximizing the spread of influence through a social network", *Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining,* ACM, pp. 137.

Merriam-Webster 2014, *Meme*. Available: http://www.merriam-webster.com/dictionary/meme [2014, 07047].

Post, J.M. 2007, *The Mind of the Terrorist: The Psychology of Terrorism from the IRA to al-Qaeda,* Palgrave Macmillan.

Quinn, B. & Branigan, T. 2014, *Malaysia Airlines loses contact with plane carrying 239 people*, The Guardian.

Quodosi, S. 2012, 15 March-last update*, Small Business Strategy :Understanding Memes* [Homepage of Benchmark], [Online]. Available: http://www.benchmarkemail.com/blogs/detail/small-business-strategy-understanding-memes [2013, 20 November].

Renando, C. 2011, *The stories we tell: Social media as narrative psychology* [Homepage of Sideway Thoughts], [Online]. Available: http://www.sidewaysthoughts.com/blog/2011/08/the-stories-we-tell-social-media-as-narrative-psychology/ [2014, 0404].

Twitter 2014*, More downloads for your app with Twitter Cards*. Available: https://dev.twitter.com/ [2014, 0407].

Twitter 2013*, GET statuses/retweeters/ids*. Available: https://dev.twitter.com/docs/api/1.1/get/statuses/retweeters/ids [2014, 0407].

Veerasamy, N. & Labuschagne, W.A. 2013a, "Establishment of Trust Factors in Social Networking Sites", *Journal of Information Warfare,* vol. 12, no. 3.

Veerasamy, N. & Labuschagne, W.A. 2013b, "Determining Trust Factors of Social Networking Sites", *Proceedings of the 12th European Conference on Information Warfare and Security*, pp. 288.

Woodworth, E. 2014, *, The 9/11 Attacks, "Keeping the Lid on the Lie": Media Response to the Growing Influence of the 9/11 Truth Movement* [Homepage of Global Research], [Online]. Available: http://www.globalresearch.ca/the-911-attacks-keeping-the-lid-on-the-lie-media-response-to-the-growing-influence-of-the-911-truth-movement/5373217 [2014, 0407].

Yirka, B. 2011, *DARPA looking to master propaganda via 'Narrative Networks'*, Phys.com.

Young, T. 2012, 15 March 2012-last update*, How to Use Memes to Build EASY Backlinks & Traffic* [Homepage of The Moz Blog], [Online]. Available: http://moz.com/blog/how-to-use-memes-to-build-easy-backlinks-traffic [2013, 20 November].

Integration of a Network Aware Traffic Generation Device Into a Computer Network Emulation Platform

Suné von Solms [1, 2] and Schalk Peach[1]
[1]Council for Scientific and Industrial Research, Pretoria, South Africa
[2]School for Electronic, Electric and Computer Engineering, North West University, Potchefstroom, South Africa
svsolms@csir.co.za
speach@csir.co.za

Abstract: Flexible, open source network emulation tools can provide network researchers with significant benefits regarding network behaviour and performance. The evaluation of these networks can benefit greatly from the integration of realistic, network aware traffic into the network emulation platform. Traffic generators are often systems that replay captured traffic packet-by-packet or generate traffic according to a specified model or preconfigured sequence. Many of these traffic generators can be easily integrated into emulation platforms, but are not responsive to network behaviour and performance as it does not adapt to current network conditions and will blindly transmit data with minimal understanding of the network and its protocols (Botta, Dainotti & Pescapè, 2010). Network aware traffic generation systems communicate to the network using various protocols before traffic generation starts, enabling the generation of realistic, network aware traffic. In this paper we describe the integration of the Common Open Research Emulator (CORE), a network emulation platform, with a network aware traffic generation system, namely BreakingPoint. The integration is possible with the implementation of CORE in FreeBSD, using the Netgraph system. This solution allows real-time bidirectional communication between the virtualised nodes in the emulation platform and the external traffic generation system. This successful integration enables researchers to assess and analyse networks with the existence of realistic, network aware traffic.

Keywords: network emulation, traffic generation, integration

1. Introduction

A difficulty researchers face when studying networks is that hardware experiments have high reliability, but are expensive to acquire, maintain and difficult to expand. In addition, the setup of all the hardware devices for every network test to be run is time consuming. On the other hand, simulation environments offer a repeatable, controllable environment that can easily be modified, but lacks realism through the use of simplified and approximated models (Wei & Mirkovic, 2006).

The natural trade-off between these two extremes is network emulation platforms. Network emulation platforms offer an acceptable amount of realism and hardware requirements, as well as multiple virtual network devices that can easily be reconfigured expanded. Therefore network emulation is an attractive and cost-effective tool for researchers interested in the behaviour and performance of networks. All virtual nodes on the network emulator can run the processes that is to run on the real hardware, as the virtual device duplicates the exact functions of the external real hardware system.

A good network emulation platform provides a range of functionalities, including the ability to create virtual replicas of servers, workstations and other network hardware to model existing or planned computer networks. The virtual replicas of network nodes enable users to create, install and run real life applications on the nodes. The applications can be accessed and used by the users, where real-life traffic generated from these applications flows through the network emulation. However, the traffic generated by the installed applications is not the only traffic present in a realistic network. Traffic is constantly generated by network appliances, background applications or by the control plane (Botta, Dainotti & Pescapè, 2010). Also included in the network is data received from outside the network, which can be traffic that originated remotely and is destined for one of the nodes in the network, misaddressed traffic, malware etc. (Douglass & Goldstein, 2004). The combination of all these functionalities ensures that the network emulator provides a realistic representation of the real-life network.

One problem continuously faced by researchers, however, is the realistic generation of network traffic. There exist a wide range of traffic generators, many providing non-accurate traffic patterns that can lead to

inaccurate test results (Botta, Dainotti & Pescapè, 2010), (Papadopoulos, 2012). The effects of non-realistic traffic generation in a network emulation can include (Douglass & Goldstein, 2004):

- experiments that may not repeatable,
- comparisons between tests that may be infeasible,
- test results that may be inaccurate.

These traffic generators may contain a wide range of features, like the replication of complex traffic models and packet captures, but are often just systems with limited awareness that will blindly transmit data in a preconfigured sequence into the network with minimal understanding of the network and its protocols. Not communicating with the network before traffic generation starts leads to a lack in network awareness. The generation of realistic network traffic that is aware of what is happening in the network is essential for the realistic simulation of a network (Venkatesh & Vahdat, 2008). Therefore, the use of an accurate traffic generator is imperative in a network emulation platform so that virtual networks can be as close to real life as possible.

We aim to assist in the generation of realistic network traffic in emulation platforms by integrating the BreakingPoint traffic generator (IXIA, 2012) into the Common Open Research Emulator (CORE) emulation platform (Ahrenholz, et al., 2008), (The Boeing Company, 2012).

CORE is an emulation platform that focuses on the accurate duplication of network devices through the construction of virtual machines. Protocols and applications intended for the real devices can be run on the virtual machines (nodes) without modification and these live networks can be connected to real-life external networks and devices. CORE is an open system, which enables us to seamlessly integrate generated traffic with the network emulation.

The BreakingPoint system focuses on the generation of network aware traffic. This system does not only generate traffic based on statistical data, but creates simulated users that generate the requested traffic as a real user would communicate to a server or other users in the network. The users simulated in BreakingPoint communicate to the network using various protocols before traffic generation starts, enabling the generation of realistic, network aware traffic.

The integration of these two systems would greatly enhance the accuracy of network testing as CORE users would be able to create realistic network traffic in a network emulation. This includes background and application traffic for the emulated nodes even if the application is not installed and running on the node. This will create a comprehensive network traffic profile that models real-life network scenarios.

The rest of the paper is organised in the following order: The next section considers related work. Section 3 describes the benefits of traffic generator integration. Section 4 describes the two integration methods that was considered while Section 5 concludes this paper.

2. Related work

2.1 CORE

CORE has its roots in the Integrated Multiprotocol Network Emulator/Simulator (IMUNES) system (Marko & Miljenko, 2004). CORE is built in a modular fashion which mainly consists of a GUI for the building of network topologies, a services layer which instantiates the virtual machines, as well as an application programming interface (API) that specifies how the different software components should interact. This modular architecture allows CORE to work with FreeBSD jails, discussed in Section 2.2, through the use of VIMAGE (Marko & Miljenko, 2004), (Zec, 2003).

The CORE system for FreeBSD was developed to use the existing operating system's virtualisation techniques to construct virtual networks. This container-based emulation system provides the high-fidelity emulation for layer 3 (network layer) and above. Real user and system applications as well as network protocols can run on the virtualised machines, or network nodes, without alteration. These processes are isolated in its own container with a virtual network stack and shared hardware resources. Links between the machines are created virtually, so that the virtual nodes and links form a complete emulated network. In addition, real

network interfaces may be added to the virtual node, which enables the connection of the node to a live external network or device.

2.2 FreeBSD Netgraph subsystem

FreeBSD jails provide virtual environments for running programs in isolation. Each jail has its own network stack and network variables, like interfaces and protocol states. A jail along with the network stack form a lightweight virtual machine capable of running user applications (The Boeing Company, 2012).

Netgraph, a graph-based networking subsystem available in FreeBSD, allows the arbitrary linking of different networking components. The Netgraph subsystem enables the arrangement of nodes into arbitral graphs. Nodes have "hooks" which are used to connect two nodes together, thus forming a graph or virtual network. These connections can be seen as virtual links in the network over which nodes communicate (The Boeing Company, 2012).

Recent additions to the FreeBSD 10.x kernel (Voras, 2013), (Admin-GvE, 2013) such as VPS1 (Ohrhallinger, 2010), bhyve (Dexter, 2013), (Grehan, 2013) and NetMap (Rizzo, 2012), (Rizzo, Guiseppe & Maffione, 2013) present new opportunities for enhancing node characteristics of container based emulator systems. These are the features which we are to use in order to integrate realistic network traffic. Through the use of FreeBSD jails and the Netgraph subsystem, CORE can provide applications running in real time on an emulated network where the hardware requirement is relatively small while a large number of virtual nodes can be instantiated (Ahrenholz, et al., 2008).

2.3 BreakingPoint traffic generator

The BreakingPoint system by Ixia (IXIA, 2012) is a network simulation device, or cyber tomography machine (CTM), which allows for detailed simulation of various aspects of network traffic, including users browsing, emailing, texting, talking, and spreading malware. This includes an extensive library of user applications, including Facebook and Youtube, which is updated and maintained on a regular basis. A Markov chain approach is used to increase the level of realism in user simulation and dynamic recreation of real world traffic flow in computer networks.

BreakingPoint defines traffic flow patterns through the use of Clients and Servers. BreakingPoint simulates clients communicating to servers using the applications and protocols specified by the researcher (IXIA, 2012). In this manner, the clients generate connections to the specified servers to establish live connections for communication. Before traffic is generated and transmitted, the clients communicate to the servers via the emulated network nodes to establish the current network conditions, thus enabling the generation of network aware traffic that responds and adapts to the changing network conditions.

The BreakingPoint device generates network aware traffic that can be specified and customised specifically for the intended purpose of the user. Individual traffic flows can be specified in terms of source and destination IP ranges, host definitions, protocol type, action parameters etc. A BreakingPoint system with a single 1GB blade can generate up to 5 million simultaneous sessions (application traffic sessions) at a rate of 500 000 sessions per second. The variety and volume of traffic that can be generated make the BreakingPoint traffic generator able to provide realistic, network aware traffic which can enhance a network emulator to provide real-world network and traffic conditions (IXIA, 2012).

3. Advantages in Integration

Although the CORE emulation platform provides virtual nodes that enable the running of real applications and real-time user interaction, it is of utmost importance to evaluate the virtual network under realistic networking conditions. One aspect relating to realistic networking conditions is that of realistic network traffic (Venkatesh & Vahdat, 2008).

The CORE emulation platform has a built in traffic generator, but the traffic flows are limited to generalised traffic models which includes:

- 10, 100, 512 kbps POISSON

- 3 second fixed BURSTS every 15 seconds
- 0-5 second random BURSTS every 5 seconds
- JITTER 10 kbps 0.05-0.15 seconds
- cloned packet capture (PCAP) files

Even though the CORE traffic generator offers a range of features, including the replication of packet captures, it is only generated to put a load on the network, not to look like realistic network traffic.

Many research has been done on the impact of non-realistic and simplified traffic models on network application and protocol behaviour (Botta, Dainotti & Pescapè, 2010), (Papadopoulos, 2012), (Venkatesh & Vahdat, 2008). Venkatesh & Vahdat (2008) proves how realistic background traffic can significantly improve the accuracy in test results as applications are sensitive to burstiness of background traffic. Due to these generalised traffic patterns available, it was thought viable and very useful to integrate BreakingPoint with CORE.

The FreeBSD jails used in CORE make the software configurable so that realistic network aware traffic can be routed to each virtualised node from BreakingPoint. The integration of network aware traffic by BreakingPoint in CORE offers the users of CORE more options. Users can now virtualise a variety of networks where background as well as application traffic specific to the network scenario can be created. In addition to the reconfigurable architecture of CORE, the CORE GUI can provide researchers a comprehensive overview of the complete network, individual nodes as well as the traffic flowing over the network.

CORE enables comprehensive traffic flow analysis and visualisation through the use of throughput and display widgets as well as packet capturing via tshark (Wireshark, 2013) on all node interfaces. The flow of realistic generated traffic as well as the traffic from the applications running on the nodes can be monitored live using the CORE GUI. CORE allows run-time interaction with the network from the virtualised nodes and external devices, which can influence application performance and behaviour. Due to the fact that BreakingPoint generates application aware traffic, it is aware of the behaviour of the network and the nodes it is communicating with. Thus, a change in the network would influence the traffic flow to that node as it would in real-life.

4. Integration and performance

As CORE and BreakingPoint are two different systems running on separate hardware, the interaction needs to be managed and coordinated. This interaction can be managed in various different ways, each with their own advantages and disadvantages. Subsequently, the two different integration methods are discussed.

4.1 Packet routing

When a CORE emulation is instantiated, each virtualised node in the emulation is able to communicate with external devices or networks in real-time. To enable communication with external devices, emulated nodes can be linked to physical networked interfaces on the host machine.

Traffic generated by BreakingPoint can be routed through an emulation by connecting a physical network interface to an emulated router. As this is network aware traffic, the traffic can be routed throughout the network and back to the BreakingPoint system through the same or another emulated router connected to BreakingPoint. A diagram explaining the traffic flow is shown in Figure 1.

As seen in Figure 1, the BreakingPoint device (Client Simulation side) is connected to the emulated network via a physical connection to the host machine. That physical network interface is linked to an emulated router, presented as the circle on the left in the diagram. The simulated clients on BreakingPoint communicate to the connected router, establishing the network configuration so that network aware traffic can be generated. It can be seen that the router forms part of the emulated network running on the host operating system. As shown by the bidirectional arrows, the traffic generated over the network can be routed back to the BreakingPoint system (Client Simulation side or Server Simulation side).

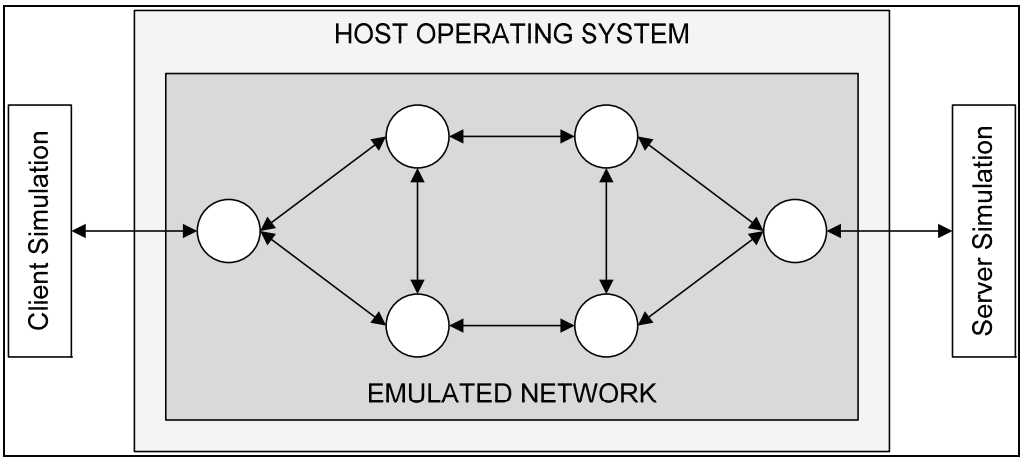

Figure 1: Integration through Packet Routing

The disadvantage associated with this configuration is that BreakingPoint can simulate application and background traffic from thousands of users, though all traffic will appear to enter the emulated network from a single point, namely the router attached to the BreakingPoint device. Thus, using this packet routing based integration shown in Figure 1, end-user devices in the emulated network will not appear to generate the externally generated traffic, as it would appear to originate from a single router. To increase the fidelity of the emulation, an outside observer must not be able to distinguish between the origin of traffic generated by BreakingPoint and traffic generated by an emulated node.

4.2 Packet forwarding

Utilising the capabilities of the BreakingPoint device to set the address ranges of the simulated clients and servers for traffic generation, combined with the capabilities of CORE to set individual IP addresses of emulated nodes, corresponding network configurations in both systems can be created.

To make it appear as though packets generated by the BreakingPoint device originate from a specific emulated node on the edge of the network, each individual packet must be forwarded to the edge node with the corresponding source IP address and then forwarded to the rest of the network. The BreakingPoint device expects all traffic generated from the client side to reach the server side, and vice versa. All traffic must therefore be sent back to the BreakingPoint device when reaching the node with the destination IP address.

Figure 2 illustrates the configuration of the emulation of client and server traffic as well as emulation of networked devices. The configuration consists of three layers, namely the externally generated traffic, components operating from within the host operating system and the emulated network. As can be seen in Figure 2, bidirectional communication is required between all components in the configuration, illustrated by the bidirectional arrows.

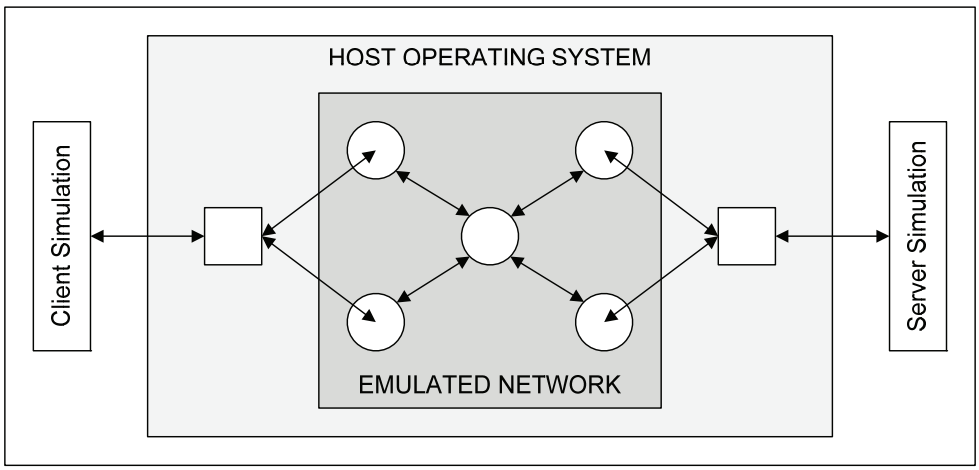

Figure 2: Integration through packet forwarding

Packets generated by the external traffic generator (BreakingPoint) is directed to physical network interface cards of the host hardware platform. A modified NetGraph bridge (ng_bridge*) node (represented by the square in Figure 2 and 3) receives all packets and forwards the packets to nodes within the emulated network (represented by circles in Figure 2 and 3). The network interface of each emulated node is a modified NetGraph Tee (ng_tee*) node, shown in Figure 3. This configuration enables all traffic received from the BreakingPoint device and all traffic generated by applications running on the node itself to be modified and forwarded into the emulated network. All network traffic coming into the node can then be modified and sent back to the BreakingPoint device, through the modified ng_bridge* node. Figure 3 shows the internal architecture of the developed mechanisms.

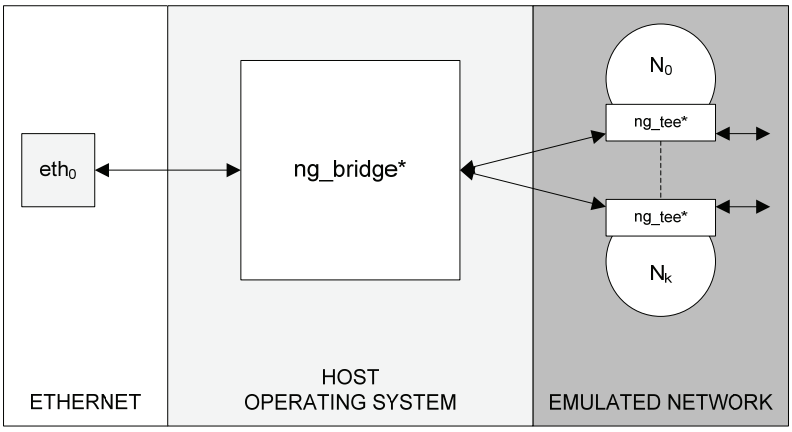

Figure 3: Packet forwarding operation

4.3 Packet forwarding operation

With both the BreakingPoint device and the CORE emulation configured to emulate the same IP ranges for both clients and servers, mechanisms to receive, modify and redistribute packets are required.

4.3.1 IP address resolution and redistribution configuration

When emulating nodes within a dynamic host address configuration environment, the integration of external traffic cannot be achieved until all emulated nodes have received an IP address. When a node receives a DHCP address from a DHCP server within the emulated network, the address assigned to the node is extracted by the modified ng_tee* node and a message containing the configuration is sent to the modified ng_bridge* node. The modified ng_bridge* updates a look-up table that associates source IP address configuration with the ng_hook* that is attached to the forward hook (see Figure 5) of the node configured with the associated IP address. Any incoming packet (generated by BreakingPoint) will be sent to the node corresponding to the IP source address of the packet by doing a lookup for the ng_hook* associated to the source IP address of the packet.

4.3.2 Packet forwarding

The unmodified NetGraph ng_tee module has four hooks, left, right, left2right and right2left (FreeBSD Man Pages, 2004). The unmodified ng_tee node will send all data received on left to the right and left2right hooks, whilst all data received on right will be sent to left and right2left. An illustration of the unmodified NetGraph ng_tee module can be seen in Figure 4.

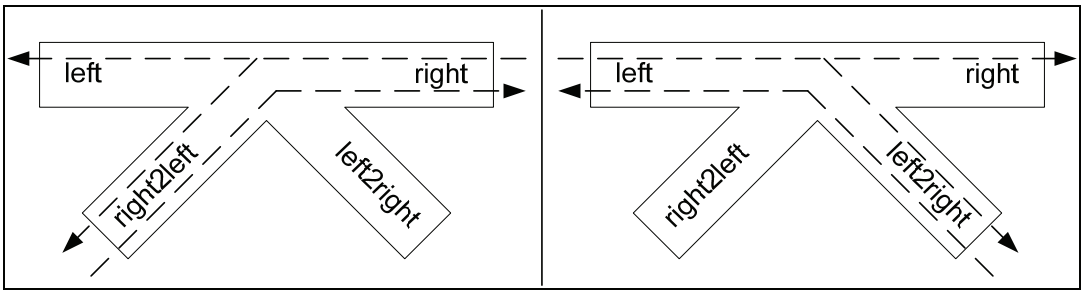

Figure 4: Unmodified ng_tee operation

For the integration of the traffic generated by BreakingPoint into the emulated nodes, the ng_tee was modified. The modified version of the ng_tee node has three hooks, inner, outer and forward. The inner and outer hooks connect the emulated node to the next emulated node in the network. The operation of the modified ng_tee* node is similar to the operation of the standard ng_tee node. All data received on inner (the emulated node itself) is sent out the outer hook (onto the next emulated node), but not to the forward hook. All data received by the outer hook is forwarded to the inner and the forward hooks. Any packet received by the forward hook is sent to the outer hook exclusively. Figure 5 illustrated the operation of the modified ng_tee* node.

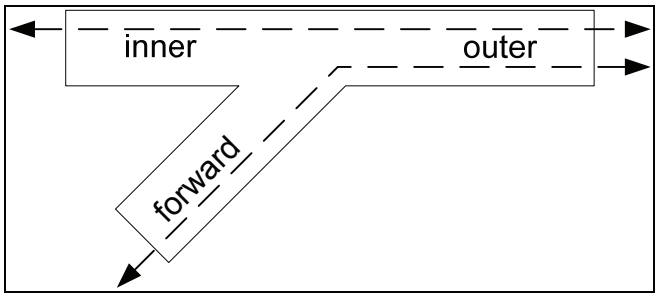

Figure 5: Modified ng_tee* operation

The forward hook receives all data sent to it by the modified ng_bridge* node, which is the data sent by the BreakingPoint device. The packets generated by the BreakingPoint device will have a MAC address that differs from the MAC address of the emulated node. Packets received by the forward hook cannot be sent on to outer hook raw, as any routing or switching nodes will then receive packets with the same IP address, but conflicting MAC addresses. In order to prevent conflicts, each packet received by the forward hook is modified to have the MAC address of the emulated node. After packet modification, a checksum recalculation is done to ensure that packets are not rejected. Before the packet can be sent to the BreakingPoint device from the outer hook, the packet is modified to have the original MAC address that was generated by the BreakingPoint device, and sent to the forward hook.

4.3.3 Complications in packet forwarding

Packet forwarding enables the integration of network traffic generated by an external device. However, due to the unconditional operation of the outer hook, the kernel of the emulated node and the BreakingPoint device will receive unexpected packets. Packets intended for processes running in the emulated node reaches the BreakingPoint device and packets intended for the BreakingPoint device will reach the kernel of the emulated node.

4.4 Discussion

This configuration shows how packets generated by BreakingPoint for a specified client can be routed to its emulated node counterpart seamlessly. With this configuration, two users communicating via video chat, for example, can be generated with the traffic flowing over the network. When an outside observer looks at the network, it would seem that the two specified nodes are generating the traffic, with the correct IPs, protocols etc. This configuration allows network researchers to create networks that mirror realistic network environments.

5. Conclusion

In this paper we discussed the integration of the Common Open Network Emulator (CORE) with the BreakingPoint device, which is a traffic generator. A brief overview was given on the two systems and we discussed why CORE is the ideal platform to use for BreakingPoint traffic integration. We motivated why such an integration will be advantageous to network researchers performing network behaviour and performance tests in the CORE emulation environment.

Through the use of the Netgraph system in the FreeBSD platform, we could successfully develop integration mechanisms enabling the flow of network aware, bidirectional traffic from each node in the emulated network. The complete technical setup in order to create Packet Forwarding is described. Future work would include the separation of the packets intended for the kernel and the BreakingPoint device.

This integration of CORE emulation platform and the BreakingPoint device enables researchers to assess and analyse networks with the existence of realistic, network aware traffic.

References

Admin-GvE, 2013. FreeBSD 10's New Technologies and Features. [Online] Available at: http://www.freebsdnews.net/2013/09/20/freebsd-10s-new-technologies-and-features/[Accessed 15 02 2014].

Ahrenholz, J., 2010. Comparison of CORE network emulation platforms. MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010, pp. 166-171.

Ahrenholz, J., Danilov, C., Henderson, T. R. & Kim, J. H., 2008. CORE: A real-time network emulator. Military Communications Conference, 2008. MILCOM 2008. IEEE, pp. 1-7.

Botta, A., Dainotti, A. & Pescapè, A., 2010. Do you trust your software-based traffic generator?. IEEE Communications Magazine, 48(9), pp. 158-165.

Dexter, M., 2013. BSD Hypervisor. [Online] Available at: https://wiki.freebsd.org/action/show/bhyve?action=show&redirect=BHyVe [Accessed 20 02 2014].

Douglass, P. R. & Goldstein, B. W., 2004. Background Internet Packet Traffic through Home Cable Connections. College of Computer and Information Science.

FreeBSD Man Pages, 2004. FreeBSD Kernel Interfaces Manual. [Online] Available at: http://www.freebsd.org/cgi/man.cgi?query=ng_tee[Accessed 20 02 2014].

Grehan, P., 2013. Extending bhyve beyond FreeBD guests. Proceedings of EuroBSDCon.

IXIA, 2012. BreakingPoint 2.0 User Guide, Release 3.0, s.l.: s.n.

Marko, Z. & Miljenko, M., 2004. Operating system support for integrated network emulation in imunes. 1st Workshop on Operating System and Architectural Support for the on demand IT InfraStructure (OASIS).

Ohrhallinger, K. P., 2010. Virtual Private Systems for FreeBSD. Proceedings of EuroBSDCon.

Papadopoulos, G. Z., 2012. Experimental Assessment of Traffic Generators, s.l.: s.n.

Rizzo, L., 2012. Netmap: A Novel Framework for Fast Packet I/O. Proceedings of the 2012 USENIX Annual Technical Conference.

Rizzo, L., Guiseppe, L. & Maffione, V., 2013. Speeding Up Packet I/O in Virtual Machines. ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS).

The Boeing Company, 2012. CORE Manual. s.l.:s.n.

Venkatesh, K. V. & Vahdat, A., 2008. Evaluating Distributed Systems: Does Background Traffic Matter?. Berkeley, CA, USA, USENIX Association, pp. 227-240.

Voras, I., 2013. What's new for FreeBSD 10. [Online] Available at: https://wiki.freebsd.org/WhatsNew/FreeBSD10 [Accessed 20 02 2014].

Wei, S. & Mirkovic, J., 2006. A realistic simulation of internet-scale events. New York, NY, USA, ACM.

Wireshark, 2013. tshark. [Online] Available at: http://www.wireshark.org/docs/man-pages/tshark.html [Accessed 20 02 2014].

Zec, M., 2003. Implementing a Clonable Network Stack in the FreeBSD Kernel. Proceedings of the 2003 USENIX Annual Technical Conference.

Legal Solutions to State-Level Cyber Intrusion Under International law: A Maze of Legal Uncertainty or not?

Murdoch Watney
University of Johannesburg, South Africa
mwatney@uj.ac.za

Abstract: One of the reasons for state-level intrusion may be attributed to a government ensuring the safety and security of its territory which includes cyberspace. The existing international law prohibits state-level cyber intrusion that constitutes an intervention, threat of or use of force or armed attack. International lawyers must interpret which cyber intrusion falls under international law. In the latter respect, the Tallinn Manual on the International Law applicable to Cyber Warfare is a valuable tool as it provides guidelines in such determination. The discussion will however show that establishing the type of intrusion is not easy as cognizance must be taken of the doctrinal law as outlined in the international law and interpreted by international lawyers and state practice which refers to how nation-states react to intrusion. The possibility of excessive state-level intrusion exists. Mass surveillance of other states may for example illustrate an imbalance between states as some states have the technical ability to dominate cyberspace to advance their own interests and motives. Finding legal solutions to state-level cyber intrusion under the present international law are uncertain. Cyberspace governance can only be achieved under the international law by the involvement of all states on an equal basis to ensure transparency and trust, with the latter being tenuous at present. A code of conduct for cyberspace may limit excessive state-level intrusion within the ambit of the international law. The multinational Convention on Cybercrime of 2001 should be re-evaluated to establish whether it will attain global recognition. In order to address the challenges in the short term a code of conduct in cyberspace should be negotiated, whilst a cybercrime treaty would be required to address challenges over the long term. State-level intrusion ultimately involves international politics and national foreign policy, both of which could be murky at times to negotiate.

Keywords: state-level cyber intrusions, international law, cyber war, Tallinn Manual, code of conduct, cyberspace governance

1. Introduction

Twenty-five years ago Berners-Lee developed the world wide web (www) to facilitate the exchange of information around the world by means of the internet. Although the web is but one part of an information and communication technology revolution, it is often described as the killer application that enabled the internet to achieve its full potential (Hall 2014). The web grew and evolved at a phenomenal rate and it changed the world profoundly, not only for the government of a nation-state, but also for its citizens. It is not an overstatement that the internet (and the www) has become an integral part of many societies. Electrical grids, air control systems and financial systems are but a few infrastructural systems that depend on information and communication technology. This dependence on the internet reveals an unfortunate vulnerability to potentially catastrophic cyber intrusions.

This vulnerability is a source of concern to many governments. Governments must ensure the safety and security of the internet. This is one of the reasons why the UK hosted an exercise on 19 March 2014, formally known as the Cyber Security Challenge in which participants competed against each other in solving a 24 hour simulated cyber attack by a foreign government on the UK's financial system (Hatfield 2014). One of the purposes of the exercise was to highlight vulnerabilities within the cyber security sector. The exercise was aimed at establishing whether sufficient technical skills exist to counter-act such an attack. This exercise illustrates the fallibility of technological security measures (Kramer in Kramer, Starr and Wentz 2009). This discussion will consider whether a state-to-state intrusion of a financial institution falls under the international law.

The above-mentioned exercise may have been met with skepticism and considered as an overstatement of the seriousness of such a threat, but Clarke and Knake (2012) warn that cyber attacks on banks could unravel the entire global financial system. Major US banks have reportedly come under attack from a foreign government (Gorman and Yadron 2013). These attacks in the form of distributed denial of service (DDoS) attacks involve multiple computers located outside the cyberspace of the US. Although it will constitute a crime in terms of the national US law, the intrusion targets the nation from outside its borders which complicates the investigation and accountability of the perpetrator(s) and in this instance, the foreign government. If an

intrusion falls under international law, a nation-state will have to deal with it within the ambit of international politics and foreign policy.

Rogers, the chairman of the US House Permanent Select Committee on Intelligence reported in July 2013 of hacking attacks allegedly launched by foreign nations such as China, Iran and Russia against the US (Kredo 2013). He also warned against the possibility of Russia or China clandestinely implanting malicious codes into the US electrical grid and activating it at a later stage. Rogers concluded that "(w)e are in cyber war today. Most Americans don't know it. They go about their lives happily" (Kredo 2013). Schmidt and Cohen (2013) agree that "(i)t is fair to say that we're living in an age of state-led cyber war, even if most of us aren't aware of it." This discussion will consider which conduct constitutes an act of cyber war under international law.

In attempting to ensure national security, the possibility exists that without restrictions a nation-state may excessively intrude the cyberspace of another. European financial institutions were shocked to discover in 2006 that in tracking terrorist funds, the US might have covertly monitored the international financial transactions of SWIFT (bank-clearing system)(Clarke and Knake 2012). The 2013 Snowden revelations of mass US surveillance against foreign governments have not only affected international relations between the US and other states, but the attention is now firmly focused on cyberspace governance in finding legal solutions to address unrestricted state-to-state intrusion. The main aim of the discussion it to establish whether legal solutions to state-level intrusion exist under the present international law. If shortcomings are identified, possible legal solutions will be suggested.

2. Understanding the motivation for state-on-state cyber intrusions

It is important to understand why a nation-state would intrude into the cyberspace of another state. Although cyberspace is not only inhabited by foreign nations, but by many other entities such as corporations or hacktivists, the discussion will focus on the conduct of a foreign state or conduct sponsored by a foreign state.

Cyber intrusion is not a legal term and in this contribution it is used within the context of a state intruding the cyberspace of another state by means of for example hacking for the purpose of espionage or leaving behind a logic bomb, theft of intellectual property, launching of a cyber attack such as a DDoS attack on the electricity grid, the defacement of a governmental website or disrupting the network. The two most discussed examples of state-level intrusions are the following:

- A DDoS attack launched in 2007 against websites of banks, government institutions and newspapers which resulted from the Estonian government's decision to remove a Russian World War II memorial in its capital, Tallinn (Clarke and Knake 2012, Schmidt and Cohen 2013).
- The Stuxnet worm used in 2010 to infiltrate the monitoring systems of Iran's Natanz nuclear-enrichment facility, causing the centrifuges to abruptly speed up or slow down to the point of self-destruction (Clarke and Knake 2012, Schmidt and Cohen 2013). The attack was aimed to prevent the possible development of nuclear weapons.

Cyber intrusion will invariably constitute a cybercrime and may fall within the ambit of the:

- National law of a nation-state where the intrusion will either affect the national security or law enforcement.
- International law which may be applicable to state-to-state intrusions in certain circumstances.

Where the intrusion was committed from outside the borders of a nation-state, investigation and accountability may be problematic, since foreign governments will have to render assistance in the investigation and to establish attribution.

A government must ensure the safety and security of its territory which includes the internet, against various threats. Much of the machinery of nation-states have shifted online with the consequence that states will defend and attack in cyberspace as well as on land, at sea, in space and in the air (Hammersley 2013). Securing a nation's cyberspace is closely linked to cyber power. Cyber power refers to the ability of using cyberspace to advance interests and goals (Kuehl in Kramer, Starr and Wentz 2009). State-level intrusion may be asymmetrical in the sense that a nation with a weak military may have the technical ability to intrude in the cyberspace of another. However, not all connected states are equal in respect of technological capabilities and a nation with advanced technological abilities may use this imbalance against foreign states to advance its

social, economic, political, diplomatic and/or military goals. Hammersley (2013) refers to the French government that targeted foreign-owned corporations in competition with French corporations. The Ukrainian and Russian conflict in early 2014 coincided with the use of cyber intrusion such as espionage and DDoS attacks, but whether it was conducted by a nation-state or state-sponsored is not clear as attribution is not easily established without conclusive evidence of state responsibility (Levintova 2014).

Doubt has been expressed whether the threat of state-to-state intrusion is so serious that it would actually escalate into cyber war under international law (Schmidt and Cohen, 2013). The possibility of such a serious intrusion nevertheless exists. Who would have foreseen the 11 September 2001 terrorist attacks on the US? In 2009 the North Atlantic Treaty Organisation Cooperative Cyber Defence Centre of Excellence (NATO CCD CoE), an international military organization based in Tallinn, Estonia took the possible threat of cyber war so seriously that it invited an International Group of Experts (hereafter referred to as the Tallinn Experts) to produce a manual on the international law governing cyber warfare. This invitation resulted in the 2013 publication of the Tallinn Manual on the International Law applicable to Cyber Warfare (referred to as the Tallinn Manual)(Schmitt 2013). The Tallinn Manual addresses both components of cyber war, namely

- *Jus ad bellum* (the law governing the right to the use of force by states as an instrument of their national policy); and
- *Jus in bello* (the law governing conduct in armed conflict and the treatment of combatants and civilians in time of armed conflict)

This contribution will primarily focus on the *jus ad bellum* component of cyber war. The rules and principles governing state-to-state intrusion under the international law will now be briefly discussed.

3. State-to-state cyber intrusion under international law

3.1 A synopsis of the rules and principles of doctrinal international law

International law is generally prohibitive in nature and therefore an intrusion that is not forbidden is presumptively legal. The international law forbids state-level cyber intrusion that constitutes:

- An intervention;
- The threat of use of force or the use of force; or
- An armed attack.

International law does not define which state-level intrusion falls within the prohibited categories or when the threshold of a category has been surpassed. International lawyers will make this determination.

Although the prohibition of intervention is not expressly set out in the United Nations Charter, the prohibition is implicit in the principle of the sovereign equality of states laid out in article 2(1) of the UN Charter. Article 2(4) of the UN Charter states: “(a)ll members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state or in any other manner inconsistent with the purposes of the United Nations.” Article 51 of the Charter provides: ”(n)othing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations, until the Security council has taken measures necessary to maintain international peace and security...”

A distinction must be drawn between the use of force in article 2(4) and armed attack in article 51 of the Charter. A state may only resort to the use of force if it was the victim of an armed attack otherwise it will violate the use of force prohibition in terms of article 2(4). If force was used against a victim state, the state will have to make use of countermeasures to ensure compliance with the international law.

3.2 Interpretation of international law

International lawyers have to assess whether an intrusion constitute an intervention, use of force, armed attack or aggression. Certain criteria may be applied on a case-by-case basis in establishing the category of state intrusion and when the threshold of a state intrusion rise from unlawful intervention to illegal use of force to armed conflict. To determine whether state-to-state cyber intrusion constitutes an intervention, use of force or armed attack, the Tallinn Manual refers to various rules. Since this discussion focuses on

determining which intrusion would fall under the international law and which intrusion may constitute an act of war, only those rules relevant to the discussion will be referred to.

3.2.1 States and cyberspace

Rule 1 provides that a state's cyberspace is sovereign. The Tallinn Experts agreed that a cyber operation directed against another state may violate the latter state's sovereignty. However, interference does not automatically violate the international law prohibition on intervention. If an intrusion is not forbidden in international law, then it is legal. For example international law does not prohibit propaganda, psychological operations, espionage or economic pressure *per se*. The Tallinn Experts were also of the opinion that the breaching of protective virtual barriers (e.g. the breaching of firewalls or the cracking of passwords) does not constitute intervention. They could not achieve consensus as to whether the placement of malware that causes no physical damage constitutes a violation of the non-intervention principle. The Tallinn Experts indicated that the non-intervention principle is violated in the instance of political interference during elections where the online services of one party are shut down or news are altered to favour a particular political party.

Rule 6 provides for legal responsibility of a state for a cyber operation that is attributable to it and which constitutes a breach of an international obligation. The rule is based on the customary international law of state responsibility which prohibits a state from inflicting significant damage on another state. In the absence of damage or where the damage cannot be attributable to a state, no responsibility attaches. The Tallinn Experts clearly stated that the conduct of private citizens undertaken on their own initiative, so called hacktivists or patriotic hackers, will be excluded from state responsibility.

In terms of Rule 9 a victim state may resort to proportionate countermeasures against the responsible state, but the majority of Tallinn Experts agreed that countermeasures may not involve threats or use of force. States facing the use of force not amounting to an armed attack will have to resort to countermeasures if they wish to respond lawfully. Dugard (2011) describes countermeasures as self-help measures not involving threats of or use of force. It must also be distinguished from retorsion. The latter is so-called unfriendly although lawful measures a state may take against another. An example of retorsion was the suspension by Estonia of some services to internet protocol (IP) addresses from Russia. It should be noted that although Iran did not publicly denounce those responsible for the Stuxnet attacks, retaliatory steps were taken: in August 2012 a virus infected the information network of the Saudi Arabian oil major, Aramco, and erased data on three-quarters of its corporate computers and in September 2012 a series of sequential attacks were launched against the US financial industry including JPMorgan and Wells Fargo which resulted in the slowing down of overwhelmed servers and denying customers access to the bank services (Chaulie 2013).

3.2.2 Ius ad bellum (right to use force)

Regarding the use of force in terms of article 2(4) of the UN Charter:

Rule 11 provides that cyber actions constitute use of force when its scale and effects are comparable to non-cyber operations rising to the level of a use of force. In determining when the intrusion has risen to the threshold of use of force, the Tallinn Experts refer to factors that may be taken into account to assess the intrusion.

Applying the scale and effects standard to Estonia, the DDoS attack was not use of force in terms of article 2(4) of the UN Charter. Funding a hacktivist group conducting cyber operations as part of an insurgency (revolution) will not be use of force, but providing an organized group with malware and the training necessary to use it to carry out a cyber attack against another state will constitute use of force. Similarly providing a sanctuary (safe haven) to those mounting cyber operations of the requisite severity will not be use of force but if the provision of sanctuary is coupled with other acts such as substantial support or providing cyber defenses for the non-state group, it could in certain circumstances amount to use of force. Activities that amount to economic or political coercion (pressure) do not constitute prohibited use of force.

Regarding an armed attack in terms of article 51 of the UN Charter:

Rule 13 states that the scale and effects required for an act to be characterized as an armed attack necessarily exceeded those qualifying as use of force. Only in the event that the use of force reached the threshold of an armed attack is a state entitled to respond by using force in self-defense. The Tallinn Experts agreed that any use of force that injures or kills persons or damage or destroys property would satisfy the scale and effects requirement for an armed attack. They agreed that the attack on Estonia was not an armed attack as the scale and effects threshold was not reached nor could the attack be attributed to another state. They also agreed that acts of cyber intelligence gathering and cyber theft as well as cyber operations that involve brief or periodic interruption of non-essential cyber services do not qualify as an armed attack.

Although the Tallinn Experts were unanimous that Stuxnet was use of force in terms of article 2(4) of the UN Charter, they were divided on whether its effects were severe enough to constitute an armed attack. Some experts were of the view that the physical damage caused to Iranian centrifuges at a nuclear fuel processing plant may have reached the armed attack threshold.

Fidler (2011) indicates that a plausible argument can be made out that the deployment of Stuxnet constituted a deliberate, hostile, highly sophisticated, state-created and critical infrastructure threatening offensive use of malware, which constituted an illegal use of force, armed attack and an act of aggression. He goes on to say that although doctrinal analysis is important, state practice must also be taken into account. Fidler remarks that nation-states have curiously been quiet about Stuxnet. He states that nation-states such as the victim state (Iran), emerging great powers not suspected of involvement (for example China, Russia and India) and developing countries have refrained from applying international law on the use of force, armed attack and aggression. He opines that the state practice of silence suggests that from a legal and technical perspective, states may not have perceived that this situation triggered the international rules on the use of force, armed attack and aggression. He concludes – and this may necessitate some debate - that after Stuxnet there may have been a development of cyber-specific rules that increase the political and legal space in which states can use cyber technologies against one another.

The Tallinn Experts could not agree on whether actions that do not result in injury, death, damage, or destruction, but would otherwise have an extensive negative affect would constitute an armed attack or not. They were for example divided in respect of a cyber incident directed at the New York Stock Exchange that resulted in the market to crash: some felt that the mere financial loss did not constitute damage for purposes of an armed conflict whereas others were of the opinion that the catastrophic affect such a crash could have constituted an armed attack. However, a cyber operation directed against major components (systems) of a state's critical infrastructure that caused severe, although not destructive effects, would qualify as an armed attack.

In respect of cyber espionage by state A against state B that unexpectedly results in significant damage to state B's critical infrastructure, the majority of experts agreed that intention is irrelevant in qualifying an operation as an armed attack and that only the scale and effects matter. Any response would however have to comply with the necessity and proportionality criteria (Rule 14) as well as imminence and immediacy (Rule 15). The majority of experts agreed that a devastating cyber operation undertaken by terrorists (non-state actors) from within state A against the critical infrastructure located in state B qualified as an armed attack by the terrorists.

The majority of the Tallinn Experts agreed that although article 51 does not provide for defensive action in anticipation of an armed attack, a state does not have to wait "idly as the enemy prepares to attack", but it may defend itself if the armed conflict is imminent (anticipatory self-defense). They agreed that although there has not yet been a reported armed conflict that can be publicly characterised as having solely been precipitated in cyberspace, cyber operations alone have the potential to cross the threshold of international armed conflict.

4. Evaluation of the effectiveness of the current international law in addressing state-to-state cyber intrusion

International law provides for state-level intrusion that is prohibited, but determination of the type of intrusion and the threshold is left to international lawyers' interpretation on a case-by-case basis. This may result in legal uncertainty as illustrated by the Tallinn Manual where some of the experts could not reach consensus on the type of intrusion and when the threshold have been surpassed.

The Tallinn Manual is the only document outlining the interpretation of the international law in respect of state-level intrusion. Although no discussion on state-level intrusion will be complete without reference to the Tallinn Manual which is a most commendable guide, it does not provide legal solutions to excessive intrusions which may not be contrary to the present international law, such as espionage and hacking.

State-level intrusion poses many challenges, such as the investigation to establish the origin of the intrusion and attribution. A country cannot determine an appropriate response if it cannot prove accountability. Interestingly enough in 2012 unnamed US government officials confirmed that the Stuxnet worm may be attributed to the US and Israeli governments (Schmidt and Cohen 2013), but most states will not acknowledge responsibility. It was suspected that Russia was responsible for the Estonian attacks, but it denied state responsibility and attribution to Russia could not be proven (Clarke and Knake 2012).

5. Possible legal solutions to address shortcomings under the present international law

The international landscape looks different in 2014 than before. The publication of the Tallinn Manual, the Snowden revelations of mass US espionage against other countries and the deployment of the Stuxnet cyber weapon and its possible repercussions have caused nations to seriously re-think the issue of cyberspace governance. Although espionage is nothing new, technology allows for excessive espionage if countries do not exercise restraint.

The issue of transparency between states and citizens may be one of many considerations regarding the discussion of possible legal solutions to state-level intrusion. The revelations by Snowden and Wikileaks illustrate that not only states may be offended by for example, state-to-state mass surveillance, but some citizens may also insist – and this is a contentious issue - that all information should be in the public domain (Leigh and Harding 2013). If state level intrusion is not subjected to checks and balances within an international framework, some citizens may feel morally obligated to expose government conduct.

In the past governments took opposing view of how cyberspace should be governed. On the one hand there was the so-called western nation-states under leadership of the US who opposed international governance in favour of bi- and multi-national agreements, whereas the so-called non-western nation-states under leadership of Russia and China proposed a central international body to regulate cyberspace governance issues (Watney 2012). Unfortunately trust, transparency and confidence between all the connected nations regarding governmental conduct in cyberspace are at its lowest ebb and cyberspace issues will have to be addressed within the framework of the United Nations.

It is against this background that the opening key note address at the RSA conference held early 2014 reflected the concerns of the international community when it called on governments to address the following four issues (Rashid 2014):

- Renouncement of cyber weapons,
- Cooperation in investigation,
- Ensuring economic activity and intellectual property rights, and
- Ensuring privacy.

The banning of cyber weapon usage was on the table as far back as 2009 when Russia and the US discussed a possible bilateral agreement, but there was a break-down in the talks (Moscaritolo 2009). One of the reasons for the break-down of the discussions may be contributed to a possible lack of trust between the two nations.

In 2011 nation-states such as Russia, China, Tajikistan and Uzbekistan called on the UN for a code of conduct to combat criminal and terrorist activities in cyberspace and nation-states to refrain from using technology to carry out attacks against nation-states (Kirk, 2012). In 2014 China renewed its call for a code of conduct (Jian 2014). The code calls for nations “not to carry out hostile activities or acts of aggression, pose threats to international peace and security or proliferate information weapons or related technologies.” The code is not without practical and political obstacles, such as what is understood with “cyber weapon.” Is it a tool that can only inflict harm or the threat of harm or does it also include so-called spyware such as Trojan Horses and other surveillance software, as they could be used to gather information in the preparation of an attack (Boulanin 2013). The code may not be without problems and although adherence to the code will be

voluntary, it may serve as a starting point for nation-states to discuss and agree on acceptable conduct in cyberspace. The code may go a long way in ensuring the co-operation between states in the investigation of intrusion, especially since the effectiveness of the CoE Cybercrime Convention of 2001 has been questioned (Carstensen, 2011). There have been calls for a global cybercrime treaty (Scholberg, 2012), but for now the immediate concerns regarding state-level intrusion necessitates serious attention.

6. Conclusion

International relations between states in respect of cyberspace governance were already strained prior to 2014. The employment of Stuxnet meant the US had crossed a Rubicon in cyberspace (Clarke and Knake) and the further allegations of mass surveillance, led to the international relations between states in this regard to be tenuous at the very best. The genie may be out of the bottle, but it is not too late to stop the escalation of state-level intrusion. In the quest of safeguarding and securing global cyberspace, not only for governments but also for its citizens, all nations should as equal partners in an open and transparent manner negotiate conduct of states in cyberspace within the framework of the UN. It may not be an easy process, but it is at least a start to return to the original purpose of the internet: an open environment built on trust.

This article is based on research supported in part by the National Research Foundation of South Africa (NRF)(UID85384). Opinions expressed are those of the author and not the NRF.

References

Boulanin, V. (2013) "May 13: Arem production goes cyber: a challenge for arms control," [online], http://www.sipri.org/media/newslettter/essay/Boulanin_May13

Clarke, A. and Knake, R. (2012) Cyber War, HarperCollins Publishers, New York, pp. 12-16, 184 – 185, 201, 213 – 215, 226, 248, 246 – 247, 292 – 296.

Chaulie, S. (2013) "Cyber Warfare is the new threat to the global order," [online], http://www.nationmultimedia.com/opinion/Cyber-warfare-is -….

Carstensen J. (2011) "European Cybercrime 10 years on – Why it's not working." available at http://www.infosecisland.com/blogview/13139-European-Cy... (2011)

Dugard, J. (2011) International Law: A South African Perspective, Juta, Cape Town, South Africa, pp. 279, 495 – 513, 519 – 525.

Fidler, D.P. (2011) "Was Stuxnet an Act of War? Decoding a Cyberattack," [online], http://ieeeexplore.ieee.org

Gorman, S, and Yadron, D. (2013) "Banks see U.S. Help on Iran Cyberattacks," [online], http://online.wsj.com/news/articles/SB10001424127887324734904578244302922317 8

Hatfield, H. (2014) "UK at risk of cyber attack in skills shortage," [online], http://news.sky.com/story/1226260/uk-at-risk-of-cyber-attack-in-skills-shortage

Hall, W. (March 16, 2014) "Happy birthday, dear killer app." Sunday Times Newspaper p 16.

Hammersley, B. (2013) Now for then: How to face the digital future without fear, Hodder and Stoughton, UK, pp.337 – 339.

Jian, S. (2014) "An International Code of Conduct for information Security" www.unidir.ch/.../a-cyber-code-of-conduct-the-best-vehicle-for-progress

Kirk J. (2012) "Russia pushes for online code of conduct at United Nations General Assembly," [online], http://www.computerworlduk.com/news/public-sector/33079...

Kramer, F.D, Starr, S.H. and Wentz, L. K. (2009) Cyberpower and National Security, Potomac Books, Washington, US, pp. 6 – 7, 37 – 40.

Kredo, A. (2013) "Mike Rogers: China, Iran and Russia launching cyber attacks against U.S," [online], http://freebeacon.com/mike-rogers-china-iran-and-russia-launchig-cyber-attacks-agai...

Leigh, D. and Harding, L. (2013) "Wikileaks," Guardian Books, London, Uk, pp. 1 – 13, 279 – 284.

Levintova, H. (2014) "Are Russia and Ukraine on the verge of an all-out-cyberwar? [online], http://www.motherjones.com/mojo/2014/03/cyber-war-ukraine-russia

Moscaritolo, A. (2009) "U.S. and Russian officials talk cyberissues," [online], http://www.scmagazine.com/us-and-russian-officials-talk-cyberissues/article/159588/

Rashid, F.Y. (2014) "Treaties, multi-national agreements needed to ban cyber wepons: RSA Chief," [online], https://www.securityweek.com/treaties-multi-national-agreements-needed-ban-cyber-...

Schmitt, M. (2013) Tallinn Manual on the International Law Applicable to Cyber Warfare, Cambridge, University Press, UK, pp. pp. 1 – 11, 15 – 18, 29 – 34, 36 – 41, 42 – 61,83 – 84, 192 – 195.

Schmidt, E. and Cohen, J. (2013) The New Digital Age, John Murray Publishers, UK, pp. 7, 103 – 120.

Scholberg S. (2012) "International Law as Framework for Peace and Security in Cyberspace," [online], www.cybercrimelaw.net

Watney M.M. (2012) "The way forward in addressing cybercrime regulation on a global level," 2012 Journal of Internet Technology and Secured Transactions (JITST) 61 – 67.

An Annotated Bibliographical Survey on Cyber Intelligence for Cyber Intelligence Officers

Cagatay Yucel and Ahmet Koltuksuz
Yasar University, Izmir, Turkey
cagatay.yucel@yasar.edu.tr
ahmet.koltuksuz@yasar.edu.tr

Abstract: Since 2008, Cyberspace is officially regarded as a distinct military domain, along with land, sea and air for many countries (Goel, 2011). Several nations are developing defensive and offensive cyber capabilities for this domain. The nature of warfare in cyberspace is radically different than the traditional warfare: the attacks are more remote, stealthy and it may be impossible to detect the identity of the adversary (Greengard, 2010). Moreover, the time parameter becomes extremely important in this domain since conducting an attack takes only seconds even if the target is at the globally farthest point possible. Therefore, the power and impacts of cyber operations are limited by fast computation and transmission skills of your front. Nations need intelligence in this relatively new domain of war in order to know the strengths and weaknesses of other nations and themselves. With cyber intelligence of high quality, nations can assess the effects of attacks; detect their vulnerabilities, therefore mitigating the risks and implement cyber security processes based on well-defined decisions (Rudner, 2013). This paper presents a literature survey on computer science methodologies that can be useful for intelligence officers working in the cyberspace. The methodologies including defensive identification methods such as incident response strategies, social network analysis (Yip et al., 2012) (Benjamin and Hsinchun, 2012), intrusion detection systems (Zaman and Karray, 2009) and anomaly detections (Chandola et al., 2009) as well as offensive methods such as disinformation, destruction of information and communication, and advance persistent threats (APTs).

Keywords: cyber intelligence, cyber warfare, espionage, cyber intelligence analysis, cyber operations

1. Introduction

The innovation in defense and weapons industry has always been the primary key for dominating the warfare. However, the digital era brought us new fronts to fight wars. Since 2008, Cyberspace is officially regarded as a distinct military domain, along with land, sea and air for many countries (Goel, 2011). And as such, nations began to develop defensive and offensive cyber capabilities for this domain.

As of May 25th 2010, USCYBERCOM - United States Cyber Command was established with the following mission statement:

> *"USCYBERCOM plans, coordinates, integrates, synchronizes, and conducts activities to: direct the operations and defense of specified Department of Defense information networks and; prepare to, and when directed, conduct full-spectrum military cyberspace operations in order to enable actions in all domains, ensure US/Allied freedom of action in cyberspace and deny the same to our adversaries." (Defense, 2010)*

This domain has entered into Russia's military strategy as well. Russian way of war includes information warfare and it follows that information warfare against Russia will be considered warfare. The 2010 Russian military doctrine calls for "prior implementation of measures of informational warfare in order to achieve political objectives without the utilization of military forces". Russia had already made the practical test for this doctrine in 2008 in the Georgian war by combining cyber and kinetic commands. Netherlands, Germany, India, Iran and South Korea are other nations building their cyber commands. Even though China had never accepted any involvement or included cyber warfare in its strategy, it is known and accepted to be one of the most powerful players in this domain. According to the security companies such as Fireeye and Mandiant, the attacks on Australia (2013), Canada (2011), Google (2010) and many other attacks are all originating from the same group in China, "The Comment Crew" (Fireeye, 2013).

However, the nature of warfare in cyberspace is radically different than the traditional warfare. The attacks are more remote, stealthy and it may be impossible to detect the identity of the adversary. The time parameter becomes extremely important. The amount of information processed in milliseconds defines the impact of the cyber front, therefore fast computation and transmission of information becomes more important in this new type of war. Another important aspect of this domain is that offense has an advantage over defense since the defender has to control the whole structure and backbone of the infrastructure while the attacker needs to

find one point of vulnerability. Likewise, an important type of attack, an advance persistent threat (APT) defines the series of repeated efforts to possess the targeted information or system. These types of attacks require advance knowledge of the targeted system and generally include social engineering techniques to achieve its goals.

As every other domain of war, this domain needs high quality intelligence for assessing the effects of attacks, mitigating the risks, and implement cyber security policies on well-defined decisions. It is very important to understand and discriminate between perpetrators to estimate the possible effects of the attack. The perpetrator may be a state sponsored organization, an autonomous robot or a hacktivist group and to differentiate between these is only achievable in the event of having high quality intelligence prior to an attack.

This study presents a literature survey on computer science methodologies that can be useful for intelligence officers working in the cyberspace. This survey provides with set of tools, algorithms and methodologies for offensive and defensive aims. This paper is organized as follows: section 2 defines the tools and algorithms for mainly on espionage; in section 3 open source intelligence methodologies are presented and social network analysis studies are discussed. Section 4 focuses on detection and prevention systems and section 5 provides set of studies about offensive technologies and strategies. Section 6 concludes the paper with a discussion on how those methodologies could be useful for intelligence officers.

2. Cyber espionage and cyber counter espionage

This part of the paper gives brief information on software and recent academic studies on espionage in computer science.

Spy-Sense: Spyware Tool for Executing Stealthy Exploits Against Sensor Networks (Giannetsos and Dimitriou, 2013).

This work demonstrates Spy-Sense, a spyware tool that allows the injection of stealthy exploits in the nodes of a sensor network. Spy-Sense is malicious software that spies on sensor node activities and relays collected information back to the adversary. Spy-sense uses Multistage Buffer Overflow Attacks (Piromsopa and Enbody, 2006) to inject the shell codes to the targeted sensor network nodes. The main enabling factor for such an attack is defined as the limited memory problems of the sensor network nodes. Assembly codes for an arbitrary network device for attacks such as manipulation, theft, id manipulation etc. are provided in the paper.

A Timing Channel Spyware for the CSMA/CA Protocol (Kiyavash et al., 2013).

This paper presents a design of a hardware level spyware for timing design that could be embedded in network devices, implementing exponential back-off algorithm (Carrier Sense Multiple access/Collision Avoidance). A methodology to queue the covert packages into the csma/ca buffer by making use of back-off times is defined. Since the implemented spyware is at the hardware level, it is impossible to detect these types of signals by studying the radio output. Theoretical proof is provided in the paper.

Self-Healing Spyware: Detection and Remediation (Ming-Wei et al., 2007).

All of the current antispyware tools are stateless to the deleted spywares in a system. However, some spywares have grown resilient against deletion and became able to self-reproduce. This paper describes how to monitor deleted malwares and remove self-healing spywares by implementing a technique called stateful threat-aware removal system (STARS). STARS works on information flows of the system and constantly monitors the flow of the resources. The proposed methodology is backed up with experimental results.

Polonium: Tera-Scale Graph Mining and Inference for Malware Detection (Chau et al.).

This paper focuses on a new radical approach for malware detection. As the malware keeps mutating and detection solutions are outpaced by the new malware produced every year, Symantec published in 2011 a new

detection algorithm that formulates the classic malware detection problem into a graph mining problem. The algorithm assigns reputation to any software that a user (anonymous peers) may encounter. The reputation values of the files and association among the machines create a bipartite machine-file graph. The key idea of the method is to determine a file's goodness by iterating through associated machines' reputations. Performed tests are over a real machine-file graph with one billion nodes and 37 billion edges and the algorithm attains %87.1 true positive percentage of detection rate.

Anti-Spyware Analysis for iOS: An Evaluation of Current Security Products Available for iOS (Oh, 2012).

This research investigates the efficiency of anti-spyware solutions for Apple mobile operating system IOS. The devices are tested using 2 undisclosed spywares that are able to provide information about calls, messages, emails and web traffic while remaining hidden from the user. All tested products need the device to be jailbroken. According to the analysis provided, all spywares tested are required physical access to install; the spywares that spy on location data are still listed on locational services of the operating system. It would be prudent to say that most security applications, including Symantec's and McAfee's software, offer no additional security than what IOS already offers on those devices.

Automated Spyware Collection and Analysis (Stamminger et al., 2009).

In this paper, a fully-automated infrastructure is developed to analyze executables from the internet. An online database of spyware-related identifiers, signature-based scanners, and behavior-based malware detection technique are used. Over 15 million URLs are visited and 35,853 executables are installed in order to be classified. The results based on a ground truth of executables are measured and quantified. This study provides a detailed overall picture for the spywares on the internet and how to analyze them.

3. Open source intelligence and social network analysis

the Downside of Open Source Intelligence, International Journal of Intelligence and Counterintelligence (Hulnick, 2002).

Open source intelligence (OSINT) as the authors stated is the bread and butter of the intelligence community. It creates the bulk data of the intelligence. Open source intelligence makes up %80 percentage of the information and these publicly available sources will likely to grow within time. However, the downfalls do exist for these sources. Sorting through huge amounts of data becomes the biggest problems for intelligence communities. Disinformation, language translation problems, steganography and the dilemma of how much OSINT should governments provide are discussed in this work while providing the discussions with real life examples.

Authoritative Sources in a Hyperlinked Environment (Kleinberg, 1999).

This paper defines the famous HITS algorithm which is an information retrieval on any hyperlinked environment. A formulation on Hubs and Authorities is provided in the form of a mutually reinforcing graph. This formulation has connections to the eigenvectors of certain matrices associated with the link graph and these connections returns additional information of the structure of the network. This paper has led conception of the state of the art PageRank algorithm which is used by many information retrieval engines. It is worthwhile to include it in this survey since many algorithms and applications are rooted from it.

Securing cyberspace: Identifying key actors in hacker communities (Benjamin and Hsinchun, 2012).

Every hacker community has its own specific properties such as hacker reputation and how to create reputation, how the hackers in the community perceive other hacker behaviors and how culture of the community affects the hacker behavior. In this study, social hacker communities are analyzed in order to develop an identifying methodology for leaders in a community. Hackhound.org and Unpack.cn are analyzed. The analysis include Ordinary Least Squares regression analysis to understand what other properties are related with fame, average message length, number of replies per thread, number of threads that a hacker involved. Total messages are the key properties to analyze.

Cyber warfare: Connecting the Dots in Cyber Intelligence (Goel, 2011).

This study is a generalization on cyber intelligence methodologies and how they could be conducted. In this wide range paper espionage and reconnaissance, propaganda and social warfare, disabling government web infrastructure, critical infrastructure attacks and recruitment of cyber intelligence are defined and explained with case studies from the past 5 years. While discussing the methodologies and technologies for data collection and analysis, the importance of human factor in the intelligence analysis is emphasized.

Partially Supervised Learning for Radical Opinion Identification in Hate Group Web Forums (Ming and Hsinchun, 2012).

Social networks are frequently used as a tool for spreading propaganda and radical opinion. However, to extract the radical opinion from a web forum, which contains large amount of unlabeled data, can be very time consuming. This study combines text parsing, information retrieval and semi-supervised learning techniques for extracting the positive examples which are forum postings that involve propaganda and radical opinion. The key idea is to extract the main radical opinion in hate group web forums by parsing through the posts on the forum by utilizing partially learning techniques. A partially learning technique is a machine learning task to find out and improve a mathematical model for the unlabeled data using the labeled and classified training data.

Structural analysis of online criminal social networks (Yip et al., 2012).

This study has its roots from the analysis done on IRC's criminal chat rooms. The very first quantitative analysis on black market chat rooms is performed on IRC. In this paper, the criminal dialogs in some carders (Credit Card Criminals) forums are analyzed using statistical analysis. Business deals and negotiations are generally done through private messages in these social networks. Constructing and analyzing the graphs constructed by the traffic of private messages result in conclusions such as small world phenomenon, assortativity coefficients, rich club phenomenon and transitivity.

Semi-automated knowledge discovery: identifying and profiling human trafficking (Poelmans et al., 2012).

This research proposes a semi-automated and human-centered knowledge discovery methodology based on formal concept analysis. The proposed method is backed up using real life case studies from the law enforcement agency in Netherlands at the department which aims for identifying and profiling human trafficking. The problem that this research tackles is the overload of textual information coming from police reports. The key idea is to use formal concept analysis which uses concept lattices to display the people found in available police reports and track the warning indicators for them. Concept lattices allow for the detection of potentially relevant links between independent observations made by different police officers, expose pertinent persons to each other, criminal networks, and the role of certain suspects in these networks.

Terrorist Threat Assessment with Formal Concept Analysis (Elzinga et al., 2010).

This research contains the same key idea with the aforementioned paper of "*Semi-automated knowledge discovery: identifying and profiling human trafficking*". The early warnings of a potential jihadist who is becoming dangerous are provided by the National Police Service Agency of Netherlands. Formal Concept Analysis is used on police reports to extract and visualize potential jihadists and Temporal Concept Analysis is used to model how the analysis is evolved over time.

4. Defensive methods, anomaly detection and IDS /IPS

Anomaly Detection: A Survey (Chandola et al., 2009).

Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior. Anomaly detection is used in many applications in computer science such as image processing, text mining, industrial damage detection, fraud detection, intrusion detection and so on. This survey provides a structured overview of the research on anomaly detection. In the survey, anomalies are distinguished and anomaly

detection techniques are classified by the intuition behind them. Each approach and the domain in which they are being used is defined, the challenges in each domain are discussed thus providing a comparative analysis.

Training and Testing Anomaly-Based Neural Network Intrusion Detection Systems (Ray, 2013).

An artificial Neural Network consists of a collection of treatments to transform a set of inputs to a set of searched outputs, through a set of simple processing units, or nodes and connections between them (PLANQUART, 2001). A neural network IDS (NN IDS) is the version of intrusion detection systems that is able to learn the new threats and store policies for those threats after a training phase. Training a NN IDS involves supplying a dataset for abnormal and normal traffic in order to letting it store the rules for recognizing abnormalities. This work presents the tests for various NN IDS while emphasizing the importance of a NN IDS to have a low convergence rate (training), high detection rate and low failure rate (errors).

Resilient Identity Crime Detection (Phua et al., 2012).

Identity fraud has become an important problem with the remarkable increase in the usage of social media in the recent years. This work proposes a novel two layered resilient mining approach to detect identity fraud detection in credit applications. The challenges in such a domain are defined as adaptivity of the fraud and the lack of quality in data. The first level of approach is Communal Detection which is addressing the problem of finding identity theft with the real data. The second one is the spike detection which tries to sort out the data for fictitious identity theft.

A survey of coordinated attacks and collaborative intrusion detection (Zhou et al., 2010).

A collaborative intrusion detection system (CIDS) is a system which monitors and analyzes multiple networks simultaneously for intrusion and attacks. A coordinated and big scale attack is harder to detect with IDSs than local attacks since IDSs only monitor a subdomain of the internet. This paper reviews contemporary research in CIDS for detecting coordinated attacks. CIDS research is classified based on two approaches; the system architecture and the alert correlation.

Technologies, Methodologies and Challenges in Network Intrusion Detection and Prevention Systems (Nicoleta, 2013).

This paper is yet another overview for network intrusion detection and prevention systems (IDPS). In this overview, the IDPSs are classified in 4 main categories: detection methodology, activity, behavior on detection and collection and analysis frequency. Misuse based detection, anomaly based detection and stateful protocol analysis detection are main classes for detection methodology. IDPSs are categorized into 4 domains according to their activities. These are network based, wireless based, network behavior based and host based. However, with the advent of the wireless network capabilities, these domains became quite intertwined. According to their behavior, IDPS are classified into two sub categories; active and passive. Finally by collection and analysis frequency, they are classified as periodical and continuous. This paper explains and gives examples to these categories with challenges that IDPS have in each category.

Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains (Eric M. Hutchins, 2011).

Advance persistent threats is a new class of threats defined as well trained adversaries running prolonged campaigns to attack highly sensitive information and systems of national security. The adversaries conducting APTs attempt to intrude until they are successful to do so. A kill chain model is a defense mechanism for preventing such attacks by Lockheed Martin Corporation. In the proposed model (described in the paper), any repetition in the trial of intrusion creates a complication for the attack since the system collects information from the attacks and thwarts the adversary by a kill chain process.

5. Offensive techniques and APTs

Identifying Tools and Technologies for Professional Offensive Cyber Operations (Prins, 2013).

The purpose of this paper is to define the tools needed for offensive cyber operations that would be performed by government authorities. This study decomposes any operation into five phases to make it consistent with NATO standard approach of operation planning. These five phases are determining the goals, selecting targets, planning, counter attack and extracting the lessons learned. Tools and resources for each phase are identified as a matrix and a canonical process model for a cyber-attack is provided in this study.

Politically Motivated Distributed Denial of Service Attacks (NAZARIO, 2009).

Distributed Denial of Service Attack is a type of DOS where many compromised coordinated machines send a flood of traffic to overload the victim's bandwidth and servers. This paper focuses on political incidents of DDOS attacks and analyzes the motivations and aims of these attacks. The studied incidents are Hainan spy plane accident in 2001, Estonia in 2007, CNN in 2008, Georgian war in 2008, Democratic voice of Burma in 2008, Russian elections in 2007, Radio Free Europe/Radio Liberty in 2008, MSK forums of Russia in 2009, Russian opposition newspaper websites in 2008, Israel-Hamas fight in 2009, Kyrgyzstan false positive incident in 2009, Kommersant newspaper in 2008, Kazakh opposition websites in 2009, Iranian elections in 2009 and South Korean-US attacks in 2009 (Nazario, 2008b, Nazario, 2008c, Nazario, 2008a, Nazario, 2007).

Theoretical Model for Creating a Nation-State Level Offensive Cyber Capability (Ottis, 2009).

This paper proposes a theoretical model that consists of three general ways to create a nation-state level capability to inflict damage on another nation-state or even non-state actors via cyber-attack. Three models are discussed; do-it-on-your-own model, volunteer forces model and mercenary approach. These approaches are compared in terms of reliability, costs and deniability of the attacks. The author emphasizes the importance of role and control of voluntary and mercenary forces that are integrated into the state's overall offensive operations unit.

Offensive Cyber Weapons: Construction, Development and Employment (Peterson, 2013).

Complexity and difficulties of attacking to an Industrial Control System (ICS) is analyzed in this work. Developing, deploying and tracking of such devices are demonstrated mostly through the Stuxnet incident in this paper. Developing and deploying of such an attack is relatively easier than maintaining communication with the deployed weapons, most likely Trojans, in order to retrain stealth. However, this paper does not address the recent studies of cryptography protocols for Supervisory Control and Data Acquisition (SCADA) networks and Distributed Control Systems (DCS). The utilization of cryptographic protocols makes some of the discussed attacks impossible to conduct. Open source literature of such devices is also emphasized in this study.

6. Discussion and conclusion

As the critical infrastructures are the assets of a nation's economy, safety and health, the governmental attacks are almost always targeted towards them. That makes the industrial control systems key targets as can be seen in the incidences of Stuxnet in Iran, attacks in Estonia and many more. The recent developments on sensor networks of critical infrastructure protection increases the importance of attacks on embedded network devices, and such attacks would be very destructive regarding the consequences of halting the services of these critical infrastructures. The studies of Giannetsos and Dimitriou in 2013 and Kiyavash et al., 2013 are relevant to such attacks as they are targeted to the sensor networks and network devices and would be of great help to the offensive units of governments.

On the other hand, most of the anti-spywares' detects signatures in the registry files, operating system files and installed programs and remove them if found. Design of the self-healing spyware in (Ming-Wei et al., 2007) creates a sticky spyware by recreating removed components of the spyware. Anti-spyware studies in the rest of the chapter 2 provide a set of methods both for mobile and autonomous systems and discuss how effective they are. The study on İOS devices (Oh, 2012) suggests that they provide no additional security. However, there should be anti-spyware designs that benefit from the contemporary malware detection systems as clients on those devices. Collective analysis and resulting databases discussed in the study of Stamminger et al. in 2009 are crucial to the critical infrastructures.

The amount of time needed to collect the open source data shortening much following the increase in the speed of the internet as well as sorting and analyzing the data became the bottleneck in the intelligence cycle. Intelligence and Law Enforcement Agencies are building their online patrol teams in order to gather information from publicly available data on a daily basis as the amount of they have to deal with reaches hundreds of gigabytes. Therefore, the studies on social networks would be in great help for them to extract the opinions in a web forum (Ming and Hsinchun, 2012), to extract the structure out of a criminal network (Yip et al., 2012) and to identify and profile online criminals (Benjamin and Hsinchun, 2012), (Poelmans et al., 2012). Although the intelligence analysis is complexity-wise too hard to be done on a computer than on a human brain, these tools should be employed to deal with such a huge amount of data.

Anomaly detection is used in many applications in computer science; such as image processing, text mining, industrial damage detection, fraud detection, intrusion detection and so on. Such methods can be employed both to protect the critical infrastructures and to detect the identity of an adversary in a network data through combining them with IDS/IPS. Collecting the intrusion data from such systems which are employed in the critical networks and mining through them to identify similarities and characterizations would be a good technique for gaining superiority over an adversary.

The study of Prins in 2013 discusses how a national cyber offensive unit should operate and breaks down operations into five phases to make it consistent with the NATO standard approach of operation planning. Tools and resources for each phase are identified for a cyber-attack while the strategy for such an offensive unit is discussed in the paper of Ottis of 2009 according to the employment of the offensive unit. To get a good understanding of the modus operandi of governments DDOS, the political incidents of DDOS attacks should be mapped over these strategies. Therefore, a detailed analysis of such incidents (NAZARIO, 2009) takes place in this annotated bibliography.

As the governments are building their cyber operation units, the complexity of a possible cyber war is increasing with the rivalry. Many of the aforementioned studies are rooted from the scientific developments which are familiar to computer scientists. Data mining, anomaly detection, network analysis, graph theory and neural networks are the main fields taking place in this front and this study presents a survey through the recent literature on these fields that an intelligence officer would need to keep up with the latest innovations in Computer Science.

Acknowledgements

Authors are grateful to the undisclosed reviewer's comments.

References

Benjamin, V. & Hsinchun, C. Securing Cyberspace: Identifying Key Actors In Hacker Communities. Intelligence And Security Informatics (Isi), 2012 Ieee International Conference On, 11-14 June 2012 2012. 24-29.

Chandola, V., Banerjee, A. & Kumar, V. 2009. Anomaly Detection: A Survey. *Acm Comput. Surv.,* 41**,** 1-58.

Chau, D., Nachenberg, C., Wilhelm, J., Wright, A. & Faloutsos, C. Polonium: Tera-Scale Graph Mining And Inference For Malware Detection. *Proceedings Of The 2011 Siam International Conference On Data Mining.*

Defense, U. S. D. O. 2010. . U.S. Cyber Command Fact Sheet.

Elzinga, P., Poelmans, J., Viaene, S., Dedene, G. & Morsing, S. Terrorist Threat Assessment With Formal Concept Analysis. Intelligence And Security Informatics (Isi), 2010 Ieee International Conference On, 23-26 May 2010 2010. 77-82.

Eric M. Hutchins, M. J. C., Rohan M. Amin 2011. Intelligence-Driven Computer Network Defense Informed By Analysis Of Adversary Campaigns And Intrusion Kill Chains. *6th Annual International Conference On Information Warfare And Security.* Washington, D.C.

Fireeye, I. 2013. Www.Fireeye.Com/Resources/Pdfs/Fireeye-Wwc-Report.Pdf.

Giannetsos, T. & Dimitriou, T. 2013. Spy-Sense: Spyware Tool For Executing Stealthy Exploits Against Sensor Networks. *Proceedings Of The 2nd Acm Workshop On Hot Topics On Wireless Network Security And Privacy.* Budapest, Hungary: Acm.

Goel, S. 2011. Cyberwarfare: Connecting The Dots In Cyber Intelligence. *Commun. Acm,* 54**,** 132-140.

Greengard, S. 2010. The New Face Of War. *Commun. Acm,* 53**,** 20-22.

Hulnick, A. S. 2002. The Downside Of Open Source Intelligence. *International Journal Of Intelligence And Counterintelligence,* 15**,** 565-579.

Kiyavash, N., Koushanfar, F., Coleman, T. P. & Rodrigues, M. 2013. A Timing Channel Spyware For The Csma/Ca Protocol. *Information Forensics And Security, Ieee Transactions On,* 8**,** 477-487.

Kleinberg, J. M. 1999. Authoritative Sources In A Hyperlinked Environment. *J. Acm,* 46**,** 604-632.

Ming-Wei, W., Yi-Min, W., Sy-Yen, K. & Yennun, H. 2007. Self-Healing Spyware: Detection, And Remediation. *Reliability, Ieee Transactions On,* 56**,** 588-596.

Ming, Y. & Hsinchun, C. Partially Supervised Learning For Radical Opinion Identification In Hate Group Web Forums. Intelligence And Security Informatics (Isi), 2012 Ieee International Conference On, 11-14 June 2012 2012. 96-101.

Nazario, J. 2007. *Estonian Ddos Attacks – A Summary To Date* [Online]. Available: Http://Www.Arbornetworks.Com/Asert/2007/05/Estonian-Ddos-Attacks-A-Summary-To-Date/ [Accessed 02 February 2014].

Nazario, J. 2008a. *Cnn Attack Summary* [Online]. Available: Http://Www.Arbornetworks.Com/Asert/2008/04/Cnn-Attack-Summary/ [Accessed 02 February 2014].

Nazario, J. 2008b. *Georgia On My Mind* [Online]. Available: Http://Www.Arbornetworks.Com/Asert/2008/07/Georgia-On-My-Mind-Political-Ddos/ [Accessed 02 February 2014].

Nazario, J. 2008c. Political Ddos: Estonia And Beyond.

Nazario, J. 2009. Politically Motivated Denial Of Service Attacks. *In:* Czosseck, C., Geers, K. (Ed.) *The Virtual Battlefield: Perspectives On Cyber Warfare.* Iospress.

Nicoleta, S. 2013. Technologies, Methodologies And Challenges In Network Intrusion Detection And Prevention Systems. *Informatica Economica,* 17**,** 144-156.

Oh, C. G. A. B. S. A. T. 2012. Anti-Spyware Analysis For Ios: An Evaluation Of Current Security Products Available For Ios. *Annual Symposium On Information Assurance & Secure Knowledge Management* [Online].

Ottis, R. 2009. Theoretical Model For Creating A Nation-State Level Offensive Cyber Capability. *Proceedings Of The 8th European Conference On Information Warfare And Security***,** 177-182.

Peterson, D. 2013. Offensive Cyber Weapons: Construction, Development, And Employment. *Journal Of Strategic Studies,* 36**,** 120-124.

Phua, C., Smith-Miles, K., Lee, V. & Gayler, R. 2012. Resilient Identity Crime Detection. *Knowledge And Data Engineering, Ieee Transactions On,* 24**,** 533-546.

Piromsopa, K. & Enbody, R. J. Buffer-Overflow Protection: The Theory. Electro/Information Technology, 2006 Ieee International Conference On, 7-10 May 2006 2006. 454-458.

Planquart, J.-P. 2001. Application Of Neural Networks To Intrusion Detection. Sans Institute 2001.

Poelmans, J., Elzinga, P., Ignatov, D. I. & Kuznetsov, S. O. 2012. Semi-Automated Knowledge Discovery: Identifying And Profiling Human Trafficking. *International Journal Of General Systems,* 41**,** 774-804.

Prins, T. G. A. R. Identifying Tools And Technologies For Professional Offensive Cyber Operations. 8th International Conference On Information Warfare And Security, 2013 Denver, Usa. P80.

Ray, L. L. 2013. Training And Testing Anomaly-Based Neural Network Intrusion Detection Systems. *International Journal Of Information Security Science* [Online], 2.

Rudner, M. 2013. Cyber-Threats To Critical National Infrastructure: An Intelligence Challenge. *International Journal Of Intelligence And Counterintelligence,* 26**,** 453-481.

Stamminger, A., Kruegel, C., Vigna, G. & Kirda, E. 2009. Automated Spyware Collection And Analysis. *In:* Samarati, P., Yung, M., Martinelli, F. & Ardagna, C. (Eds.) *Information Security.* Springer Berlin Heidelberg.

Yip, M., Shadbolt, N. & Webber, C. 2012. Structural Analysis Of Online Criminal Social Networks. *2012 Ieee International Conference On Intelligence And Security Informatics (Isi).*

Zaman, S. & Karray, F. Tcp/Ip Model And Intrusion Detection Systems. Advanced Information Networking And Applications Workshops, 2009. Waina '09. International Conference On, 26-29 May 2009 2009. 90-96.

Zhou, C. V., Leckie, C. & Karunasekera, S. 2010. A Survey Of Coordinated Attacks And Collaborative Intrusion Detection. *Computers & Security,* 29**,** 124-140.

PhD Research Papers

Intrusion Detection System Using Bayesian Network Modeling

Chaminda Alocious, Nasser Abouzakhar, Hannan Xiao and Bruce Christianson
University of Hertfordshire, Hatfield, UK
c.alocious@herts.ac.uk
n.abouzakhar@herts.ac.uk
h.xiao@herts.ac.uk
b.christianson@herts.ac.uk

Abstract: Computer Network Security has become a critical and important issue due to ever increasing cyber-crimes. Cyber-crimes are spanning from simple piracy crimes to information theft in international terrorism. Defence security agencies and other militarily related organizations are highly concerned about the confidentiality and access control of the stored data. Therefore, it is really important to investigate on Intrusion Detection System (IDS) to detect and prevent cyber-crimes to protect these systems. This research proposes a novel distributed IDS to detect and prevent attacks such as denial service, probes, user to root and remote to user attacks. In this work, we propose an IDS based on Bayesian network classification modelling technique. Bayesian networks are popular for adaptive learning, modelling diversity network traffic data for meaningful classification details. The proposed model has an anomaly based IDS with an adaptive learning process. Therefore, Bayesian networks have been applied to build a robust and accurate IDS. The proposed IDS has been evaluated against the KDD DAPRA dataset which was designed for network IDS evaluation. The research methodology consists of four different Bayesian networks as classification models, where each of these classifier models are interconnected and communicated to predict on incoming network traffic data. Each designed Bayesian network model is capable of detecting a major category of attack such as denial of service (DoS). However, all four Bayesian networks work together to pass the information of the classification model to calibrate the IDS system. Bayesian. The proposed IDS shows the ability of detecting novel attacks by continuing learning with different datasets. The testing dataset constructed by sampling the original KDD dataset to contain balance number of attacks and normal connections. The experiments show that the proposed system is effective in detecting attacks in the test dataset and is highly accurate in detecting all major attacks recorded in DARPA dataset. The novelty of this IDS is the adaptability for the different attacks domains and the ability to update the IDS with new models. Furthermore, the practical implementation of the proposed IDS system can be utilized to train and detect attacks in live network traffic.

Keywords: cybercrimes, Bayesian network, intrusion detection system, DARPRA, adaptive learning

1. Introduction

Network and system security provides ability to computer systems to merge and work together in a more secured and trusted environment. Modern day's people used to do their online shopping, learning and bank transactions. These data transferred between different parties must not be revealed to the unauthorized third parties. Therefore, providing integrity, confidentiality and availability of the information carried through the networks is a fundamental responsibility of network security. Therefore, cybercrime protection and network security is becoming increasingly more valuable. Nowadays, networks are built with large, complex systems based upon technologies that designed without security in mind. Simultaneously, these systems are under the experiments of cryptanalysis. Cryptanalysis has made significant inroads towards vulnerable computer system and which most of them have successfully exploited. We are living in an era of nuclear wars and unstable political environment which implies that highly dangerous systems must be highly secured. The implication of cybercrimes has increased ability of unauthorized access to the systems using modern cryptanalysis automated tools. These tools have been made an attacker's job much easier and accurate.

Distributed Denial of Service (DDoS) attacks has become a severe threat to Internet security. DDoS attacks carried out to make system inaccessible by flooding the server's network and end user systems with fake generated traffic. This will stop legitimate users gain access to the system resources. Moreover, problems in network security come in the form of hostile/unwanted trespass either by users or software compromise to unauthorized login, unauthorized elevation of privilege and deployment of viruses. There are three forms of intruders identified in the field of network security; masquerade which is an outsider typical action would be unauthorized access by penetrates system or access controls to exploit a legitimate user's account. The second type of intruder is misfeasor (generally an insider), which is a legitimate user that access data programs resources for which either she/he is unauthorized to access or alternatively to misuse that he/she is authorized to access. Clandestine user (insider or outsider) typically obtains administrator privilege control to avoid auditing access controls or to suppress audit collection. This research considers detecting these types of

attacks with an anomaly based IDS (Khor at el, 2009). This research proposes an approach to build a distributed Intrusion Detection System (IDS) with the machine learning technique called Bayesian network (BN). A Bayesian network is an accurate and robust technique to build classification models to classify the normal and abnormal behaviors in network traffic data. This research has investigated towards the artificial intelligence and machine learning technique to construct anomaly based IDS with an adaptive learning process. Bayesian network provides the adaptive nature to the IDS, Bayesian networks are useful tool for determining suspicious activity or security threats in computer systems (Abouzakhar et al, 2010).

2. Research background

Most of the modern computer systems are distributed systems, such systems are more vulnerable to malicious activities rather a centralized system. The distributed nature of the system allows attackers to use more resources for the attack than centralized systems, also it is difficult to prevent or detect such attacks due to the complexity of the system [6]. In the context of network security intrusions defined as any malicious activities that could compromise the system integrity, confidentiality, or availability (Debar et al, 2000). Detecting threats using machine learning techniques are more important and one of the most reliable technique is "Bayesian network" (BN) based anomaly based detection. This will eliminate the problem of changing attack signature over the time. A Bayesian network is recognized as a tool for modelling decision equations in graphically, continuing uncertainty. The IDS model proposed by in (Abouzakhar et al, 2011) was able to build an adaptive IDS model using Bayesian networks. In this paper (Abouzakhar et al, 2011) a BN is used to build automatic intrusion detection system based on anomaly detection. In this research (Abouzakhar et al, 2011) authors presenting an adaptive probabilistic approach for intrusion detection using Bayesian network in distributed systems. In this research, Bayesian learning approach for detecting cybercrime is based on detecting signature based threats in a large distributed system dataset. In (Abouzakhar et al, 2011) the developed Bayesian consists with a learning model that corresponds to the adaptive knowledge also includes the dynamic organization of the learning detector and parent variable discovery.

The paper presented in (Howard et al, 2009) has discussed two practical real time applications of Bayesian network for distributed system intrusion detection. Voice-over-IP (VoIP) System and E-commerce system were tested based on their Bayesian network model. The specialty of this effort is Bayesian network was manually created by the expert knowledge and based on attack graph. The Bayesian networks were based on attack graphs that include several multi-step attacks for software vulnerability (Howard et al, 2009). The model explains in this research contains an attack graph, Bayesian algorithm, Inference algorithm and a controller. Bayesian network feature selection has been evaluated by the research effort in (Khor et al, 2009). There are many features selection algorithms that can use in building Bayesian network. This research proposes two feature selecting algorithms, which are used to filter important features from original dataset that can use to build the Bayesian network. Built classifier models evaluated based on three types of a Bayesian network. Firstly, naive Bayes classifiers (NBC) which involves no learning process. Secondly, the Learned Bayesian Network classifiers which uses a learning algorithm to learn the Bayesian structure. Finally, the Expert-elicited Bayesian Network normally utilizes a standard network intrusion dataset [3] to train the structure. The data set contains 4 different types of instances that fall into the category of most known attacks. Denial of service, Probing, Remote to Local (R2L) and User to Root (U2R) are the four main categories of attacks that appear in the dataset (Khor et al, 2009).

Research work presented in (Tang et al, 2009) has addressed the situational awareness, the adaptive detecting ability in the context of Insider cyber threat. Tthis approach using Dynamic Bayesian Network (DBN) as the concept of adapting the changes of the threats. In (Tang et al, 2009) authors claiming that most of traditional IDS or Intrusion prevention systems fail to address the dynamic inference capability and dynamic nature of the threats. In (Tang et al, 2009) authors proposed improved algorithm based on a DBN to recognize the dynamic behavior in the transition relationship of the Hidden Markov Model (HMM). DBN used to detect the user's session behavior dynamically and must capture these changing situations and capture the future actions based on historic data. The framework in (Tang et al, 2009) established behavior representation by two types of variables called randomly observable state variables and random hidden state variables. This representation (DBN) typically a classification model is transformed into an HMM with the application of inference algorithm. DBN to HMM transformation benefit all three stages of classification model construction, learning and inference. Grid computing is the future of distributed computer systems and grid computer systems uses proxy for credentials authentication and authorization. These credentials will be attacked even they have very short

lifetime, but system proposed in (Kunz et al, 2011) presenting a Bayesian network based classifier on this grid computing infrastructure. This Bayesian classifier will detect abuse grid credentials and legitimacy user alteration. Grid computer infrastructure slightly changed to add a Bayesian network classifier that detecting abused grid credentials details. This paper (Kunz et al, 2011) proposed an improved middleware security in the grid infrastructure that increases the grid security up to 99.5%. In this case, Bayesian classifier train data that's collected upon proxy credentials audit records that has labelled as legitimate or abuse. Bayesian classifiers build on the knowledge of both Grid path graph and the credentials graph interconnected or overlay two graphs to correlate the information in both graphs (Kunz et al, 2011).

The paper in (Frigault, 2008) has proposed a DBN approach to detect vulnerabilities in the computer system and address the issue of the changing nature of the threat by the time. When more technical details of the vulnerability published exploitability of the thread become more sophisticated. Mainly, the research (Frigault, 2008) has focused on building a DBN based IDS. The proposed DBN uses for detecting and change the model based on the evolving nature of the threat. DBN can be derived from the attack graphs and attack graphs update with updating constantly changing security environment. The model uses standard CVSS scores which ensure the model can lead to actionable knowledge (Frigault ,2008). This type of IDS endures the nature of the system and attack environment changes. Alma Cemerlic, Li Yang, Joseph M. Kizza in (Cemerlic, 2011) has presented a method with adaptive network intrusion detection with Bayesian network as the model construction technique, Bayesian training done through a mixed dataset containing real-world data and also DARPA dataset traffic. Most of research that focuses on anomaly detection using a statistically based approach, but as in (Cemerlic, 2011) authors mentioned that statistic is not based on a model of adaptive intelligent learning. Bayesian network solves the problem of adaptive learning from the past data. This proposed model that using K2 training algorithm as the Bayesian network learning algorithm and the junction tree as an inference algorithm.

3. Intrusion detection and machine learning

In this section machine learning based ID systems has been discussed in detail with the diversity of methods and modern trends of artificial intelligence and data mining techniques. ID systems were developed to detect and prevent malicious activities and provide strong security for the security critical computer systems. Modern research community investigating towards the intelligent based IDS systems that can detect intrusions with smart behaviors. ID systems are performing the detection of internal and external attackers, one of the major tasks of intrusion detection identifies the unusual patterns in user activities; provide early alarms and taking action against attacks. IDS can be divided into two major categories. Host based intrusion detection systems (HIDS) and Network based intrusion detection (NIDS). HIDS consider the data of local audit records such as windows audit files or process information, but NIDS system analyzes the network traffic data which is a recorded “tcpdump” file from a network. Signature based detection also called as rule based detection determines the user behavior with a comparison of some rules defined related to legitimize the user’s behavior. Statistical anomaly Detection is another method used in IDS systems for detecting by data collection carried out for legitimate user over a period of time. Then these learned structures (profiles) can be used to determine unauthorized users' behaviors with the highest level of confidentiality (Mukhopadhyay et al, 2011).

3.1 Signature-based intrusion detection

Signature based IDS consists with a database of known signatures of known attacks, these attacks are predefined based on the attack analysis. Most of the signature based system has a low false positive rate, which is really accurate in terms of detecting known attacks by comparing incoming network data with signature database. Modern cyber-attacks are growing rapidly and change its formation regularly. Due to this reason signature based IDS systems are difficult to maintain the demand.

3.1.1 Genetic algorithms

A genetic algorithm is a computational model, the basic concepts behind genetic algorithm is an evaluation and natural selection. This means only fittest will be survived in the process of natural selection. Genetic algorithms creating set of rules for network data with the following sample structure (Pr-owl). The below connection structure will be searched in the incoming network traffic to find out the attack connections.

If { The connection has following information: source IP address 124.12.5.18;

Destination IP address: 130.18.206.55; Destination port number: 21; Connection time: 10.1 seconds}

Then {Stop the connection}

3.2 Anomaly-based intrusion detection

This is the most common and useful IDS method used in network security. This type of methods creates profiles for the normal behavior of the network connections. This normal profiles learned IDS can be used to distinguish the incoming connections as normal or anomaly. Anomaly based ID system's major strength is reliable detection of unknown attacks. However, same time it gives so many false alarms, which eventually cause flooding out the network. The proposed IDS belongs to an anomaly based category. Machine learning techniques for Intrusion detection has become a crucial research area due to its high importance. Machine Learning based IDS systems are still in the early stages, however these methods getting popular. Researches consider the artificial intelligent (AI) techniques and data mining approaches to build more robust, accurate IDS (Kandeeban et al, 2011). In (Mukhopadhyay at al, 2011) explains most of the modern intrusion detection techniques.

3.2.1 Artificial neural networks

Artificial Neural Networks can be train with the network traffic data, then use these neural networks to recognize the patterns in network data. These neural patterns can be used to verify and distinguish between intrusion and normal connections. An artificial neural network consists with set of treatments to transform inputs to a set of searching outputs, through a set of simple processing units or nodes and connections between them (Sans Institute, 2011).. The neural network based intrusion detection uses two types of training algorithms called supervised learning and unsupervised learning. Supervised learning state is learning the desired output for a given input. Multilevel perception (MLP), this is the most commonly used supervised learning algorithms (Sans Institute, 2011). Unsupervised neural network learning algorithm (ex: self organized maps) learns without specifying the desired output (Sans Institute, 2011).

3.2.2 Bayesian network

Bayesian networks apply to many domains such as medical diagnosis, machine learning, speech recognition, signal processing, Bioinformatics, natural language processing and cellular networks (Pr-owl, 2011). Bayesian Networks are very attractive machine learning technique that represents the domain knowledge and domain information in an elegant mathematical structure with simplified visual representation. Bayesian networks are graphically represented models that show a probabilistic relationship between a set of variables under the domain of uncertainty. The Bayesian network structure is represented in a directed acyclic graph and conditional probability tables (CPTs). The CPT table represents the probability of a random variable where, given the occurrence of its parent nodes. However, can we apply the same conceptual strategy to network security with Bayesian network. The computer security and thread models are changing with the time, adaptive Bayesian network is a strategic solution for network intrusion detection.

4. Proposed IDS model

The proposed IDS has three major functionalities which implemented using WEKA Java API for machine learning. Firstly, the IDS consists with a dataset pre-processing technique such as, attribute selection, attribute filtering, and instance filtering. Secondly, the IDS consists with a Bayesian network classification model, which is the key component in the system, which does the classification of the network data. Thirdly, Inference analyzer which has designed as the prediction module for incoming testing network traffic. These modules named as data preprocessing, Bayesian network structure learning and inference algorithm module to classify the incoming new data respectively. The novelty of this IDS can express as the adaptability for the different attacks domains and the ability to update the IDS with new models.

4.1 Proposed IDS conceptual model

The proposed IDS is not merely to detect only denial of service attacks, IDS also capable of detecting other types of attacks which has appeared in testing dataset. This case, those attack types were “probes”, “user to root” and “remote to user” attacks. The major component of the system is the Bayesian network classification model which has four different Bayesian network models, which coordinates together to provide accurate detection results. This will guarantee that very less number of attacks will by-pass the IDS, beacuse each network is a specialist model to detect the special kind of attack type. The main novelty of the proposed IDS is

in the form of ability to adapt based on the input dataset. The proposed IDS is capable to work by adding more and more Bayesian network models based on different attacks. Two other major components are data distributor and Inference analyzer.

Pre-processing: Pre-processing is responsible for data preparation for the Bayesian network model learning process. This module uses attribute selection, attribute filtering and instances filtering for preparing the input dataset.

Bayesian network structure learning (Building Bayesian Models): This is the key module of the proposed IDS, proposed architecture contains four Bayesian networks to detect four different attack types. Data distributor is used to feed relevant data (network traffic) to relevant Bayesian network model for training. These models are adaptable to detect new attacks since proposed IDS support adding new Bayesian networks or modifying of existing network with attack's features.

Inference Analyser: Once all the Bayesian network models are built (trained on network traffic) and those networks are ready for predicting attacks in incoming network traffic. Test data divided by inferential analyses to each Bayesian network to classify in the attacks. Inference Analyser classify each record of the input data to normal connection or to a relevant attack. The proposed IDS model has been demonstrated below in detail.

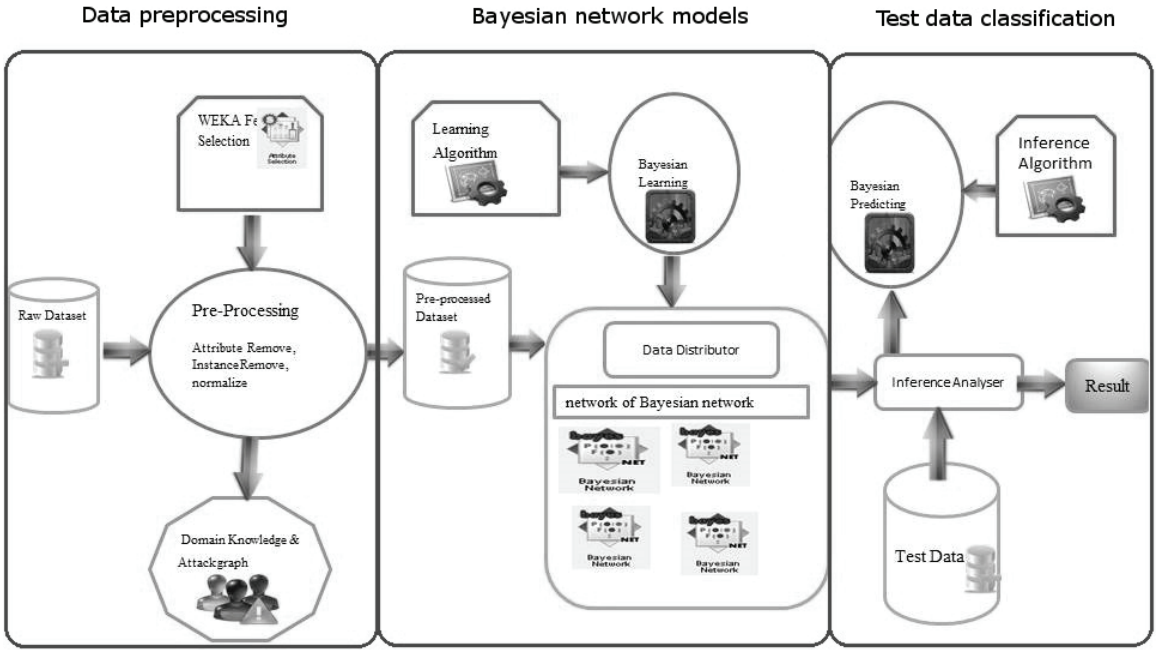

Figure 1: Proposed IDS model using Bayesian network

There are several approaches to detect intrusion using anomaly based detection, However our research is unique by the dynamic ability to construct relevant Bayesian networks for different attack domains. Data distributor is a unique concept to make the IDS much more configurable and also to plug new Bayesian network models to the system easily. These added Bayesian networks are interconnected to carry out the detection.

4.2 Feature extraction and data pre-processing

Data pre-processing is required in order to remove unwanted attributes from the original dataset and construct a dataset which can be used to build a Bayesian network. The feature selection specifically achieved by analyzing, packet information, in this section dataset feature will be analyzed based on the attacks nature and additional domain knowledge. However, since vast numbers of cyber-attacks recorded in KDD dataset, it is better to build a Bayesian network by focussing on each attack type. Identified attack types are DDOS, R2L, U2R and Probes. Data pre-processing including extracting features from original dataset. WEKA attribute filtering has been used with other pre-processing techniques such as attribute selection, attribute discreatize and filling missing instances.

Table 1: Attribute selected for each attack type

Dataset	Attributes
DDoS	This attack type has carefully studied and attributes related to this has been selected. Ex : Protocol type, service, server related attributes (server error rate), file creation, guest login, destination host error rate, destination server error rate and same service rate. (Bayesian network displays the full set of attributes).
Probes	Service, Protocol type, flag ,serror_rate , rerror_rate, same_srv_rate,diff _srv_rate , dst_host_diff_srv_rate, dst_host_srv_diff_host_rate, dst_host_srvserror_rate, dst_host_error_rate, dst_host_srv_rerror_rate
User to Root	Duration, Protocol type, service, src bytes, num compromised, hot, root_shell, dst_host_count ,dst_host_srv count ,dst_host_same_src_port rate, num_failed_logins
Remote to User	duration ,protocol_type ,service ,src_bytes, urgent ,hot ,num_failed_logins lnum_compromised, lroot_shell, is_host_login , dst_host_same_src_port_rate dst_host_rerror_rate ,dst_host_srv_rerror_rate

4.3 Data distributer

This module's major function is to distribute the training data among the separate Bayesian networks , where each category is designed to train a special kind of attacks. Data distributor is generating four separate "arff" data files that can be used to train the each different network. Input dataset has 41 attributes which must be distributed to each sub dataset based on the configuration setup by attributes removal algorithm.

4.4 Bayesian network classifiers construction

Bayesian network classifiers are built based on the training data provided by data distributor. Bayesian network classifiers building process includes structure learning, parameter learning, and building probability distribution tables for each node in the network. There are two major learning processes for Bayesian network structure learning. The structured learning, also known as casual discovery which is the process that Bayesian network used to learn the structure and parameters with the provided input data. The causal discovery aims to learn the structure and learn the parameters. The proposed IDS is using various algorithms such as K2, Hill climbing and tabusearch. The probability distribution learning achieved with algorithms such as Bayesnet estimator, BMA estimator and multinomial estimator. Once structure learning is complete then parameter learning completes the CPT tables for each feature in the BN.

4.4.1 Bayesian network design for DDOS attacks

The following network design diagram is for detecting distributed denial service attacks. Bayesian network design needs to consider the attributes, search algorithm and estimation algorithms. The "Hillclimber" search algorithm with five parents used as the search algorithm for this network with simple estimator as an estimation algorithm with threshold value "0.5". Finally, After studying the attack graphs and also with the domain expert knowledge and analysis using tools following set of features were selected to construct the DDOS Bayesian network.

The same procedure will be followed to create the Bayesian networks for other attacks types. The proposed IDS allows to integrate new Bayesian network models to the IDS based on the expert knowledge on the modern attacks. In this scenario we have designed three other Bayesian networks for attack types. Such as probing attacks , user to root attacks and remote to user attacks.

4.4.2 Inference analyser

The inference analyzers perform the classification of new incoming data with the every relevant Bayesian network model created. Inferential analysis determines which network gives the best result for particular inference effort. Attributes distribution in inferential analyses takes the new connection data and distribute the relevant attributes to the Bayesian network for inference. Once all the virtual datasets created its feed into the inference algorithm which can be used to do the classification of each dataset.

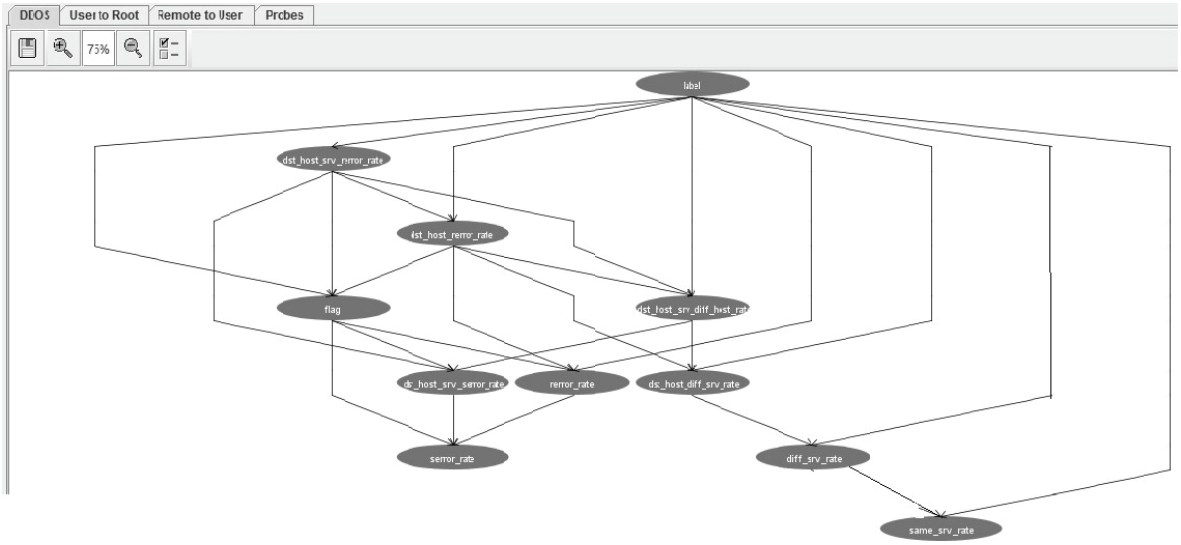

Figure 2: Proposed Bayesian network for DDOS

5. Experiments results and analysis

- Bayesian Network Learning / Training Process

The proposed IDS has been implemented using the Java WEKA library for data mining and machine learning. The implementation consists with four major module's data pre-processing, data distributor, Bayesian structure learning and Bayesian inference analysis. The proposed IDS system prototype was developed as below diagram. Developed system learned with HillClimber algorithm and generate four Bayesian networks and then use those networks to predict on a given test data file.

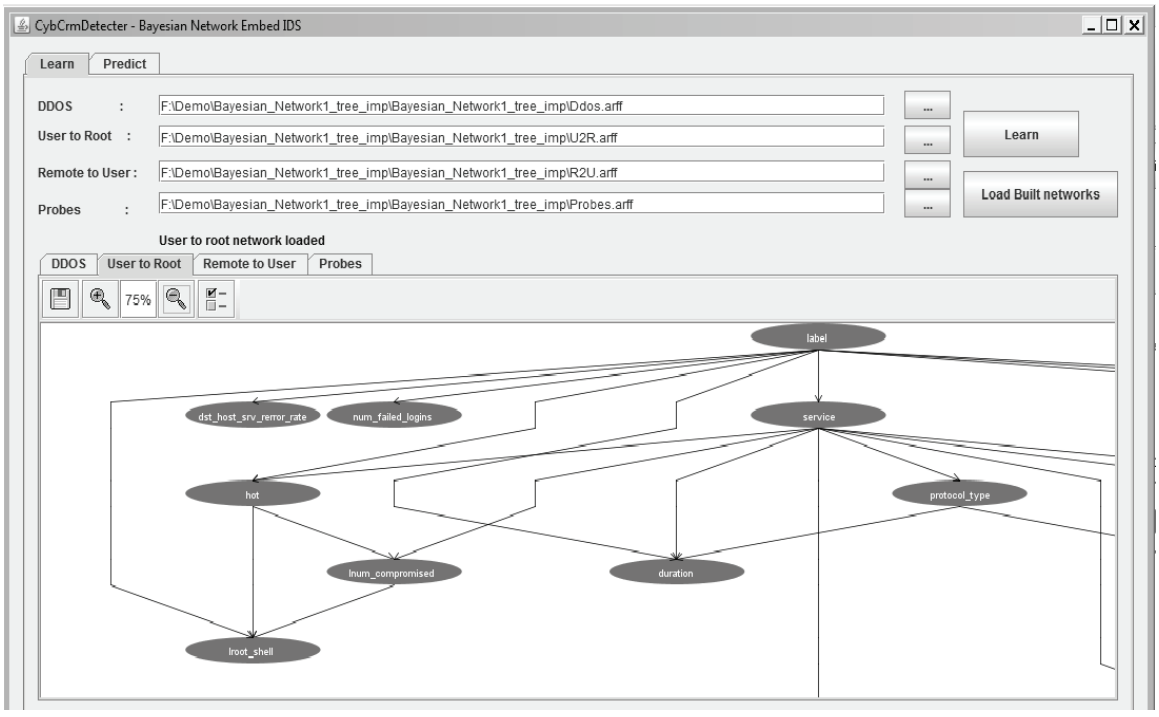

Figure 3: Proposed IDS learning stage screen

- Bayesian Network Predicting process

Prediction on new test data file and writing the result to the user is displayed as following diagram.

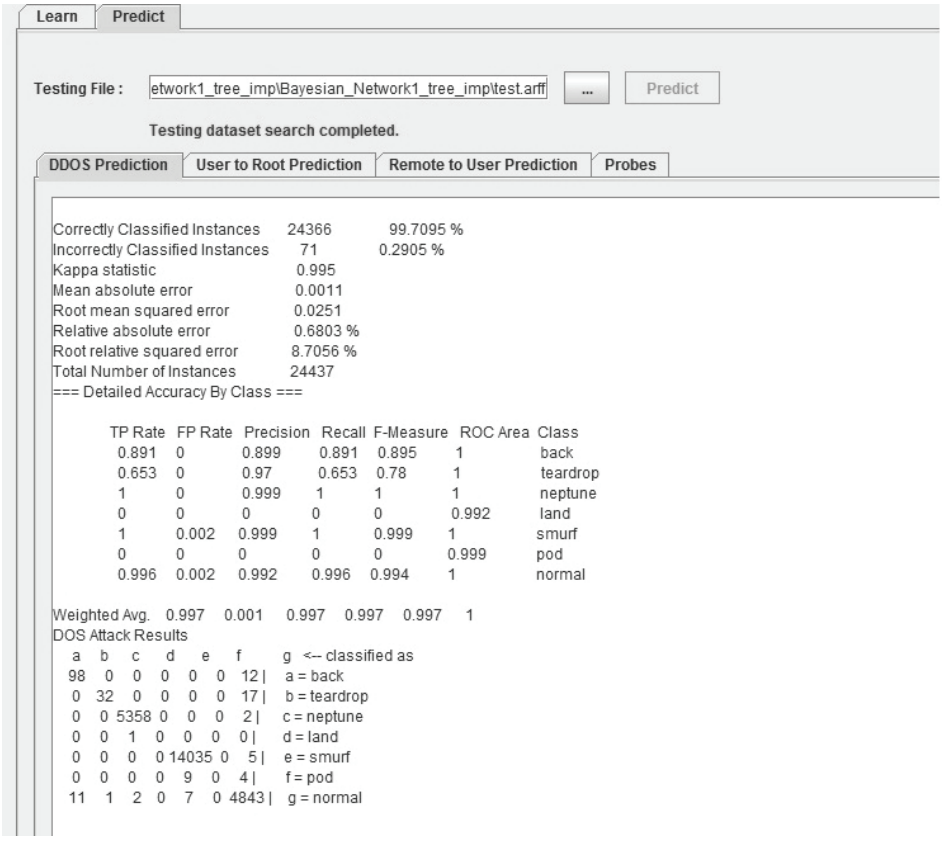
Learn
Predict
Testing File :
etwork1_tree_imp\Bayesian_Network1_tree_imp\test.arff
...
Predict
Testing dataset search completed.
DDOS Prediction
User to Root Prediction
Remote to User Prediction
Probes
Correctly Classified Instances 24366 99.7095 %
Incorrectly Classified Instances 71 0.2905 %
Kappa statistic 0.995
Mean absolute error 0.0011
Root mean squared error 0.0251
Relative absolute error 0.6803 %
Root relative squared error 8.7056 %
Total Number of Instances 24437
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
0.891 0 0.899 0.891 0.895 1 back
0.653 0 0.97 0.653 0.78 1 teardrop
1 0 0.999 1 1 1 neptune
0 0 0 0 0 0.992 land
1 0.002 0.999 1 0.999 1 smurf
0 0 0 0 0 0.999 pod
0.996 0.002 0.992 0.996 0.994 1 normal
Weighted Avg. 0.997 0.001 0.997 0.997 0.997 1
DOS Attack Results
a b c d e f g <-- classified as
98 0 0 0 0 0 12 | a = back
0 32 0 0 0 0 17 | b = teardrop
0 0 5358 0 0 0 2 | c = neptune
0 0 1 0 0 0 0 | d = land
0 0 0 0 14035 0 5 | e = smurf
0 0 0 0 9 0 4 | f = pod
11 1 2 0 7 0 4843 | g = normal

Figure 5: Proposed IDS predicting screen

Experiment of the research has obtained valuable result for detecting DDoS attacks. Experimental setup for the research presented with detail using WEKA Knowledge flow. IDS has evaluated with its true positive, false positive rates with some other parameters such as precision and recall.

5.1 Experimental setup

In the experiment each Bayesian network trains with relevant attack category instances from the input datasets and build all BN at training stage. Generated BN models are used to classify using one large testing datasets. Normalization of training datasets and testing dataset is done through the WEKA dataset unsupervised learning. Experiment setup can be modelled using the WEKA experimental designing tool as follows.

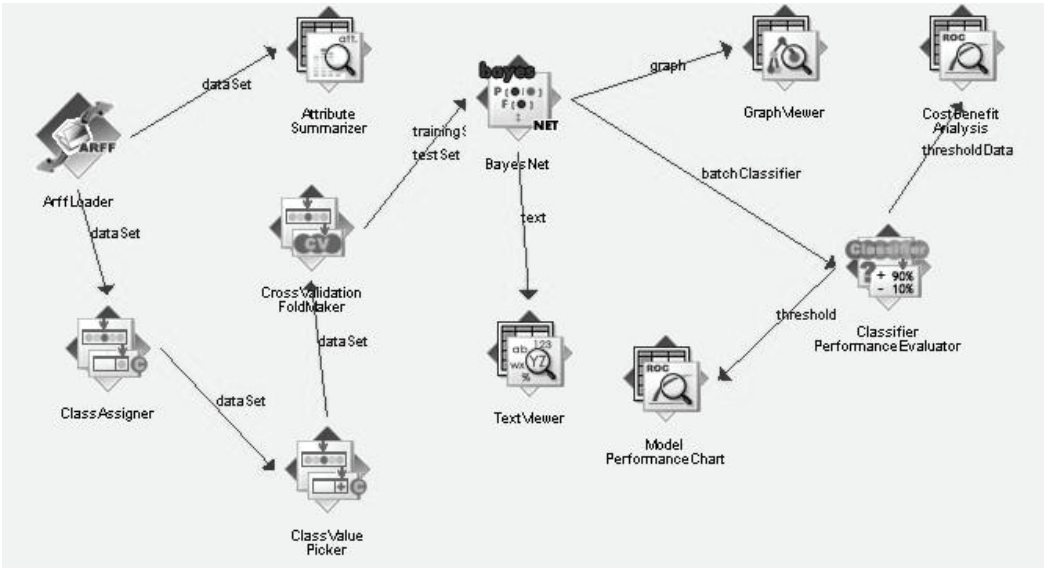

Figure 6: Model experimental setup

5.2 Result evaluation

Experimental data was captured with each attack and their true positive, false positive, precision, recall and F-Measure. Experiment data (training and testing) only based on DARPRA dataset which is quite different than the real world connection data. Therefore IDS accuracy is recorded as high it can be reduced if the experiment's data was real time.

True Positive: This is the rate of correctly classified records out of all records in the give testing dataset.

False Positive: This is the ratio of number of incorrectly classified normal connections.

5.2.1 Performance evaluation of DDOS attacks

DDOS attacks are most common and devastating attacks for security critical systems. Experiments carried out to evaluate the proposed systems DDOS attack detection ability. An experiment carried out with 10 folds cross validation, evaluation methods for the following dataset configuration for each attack. The result has categorized as dataset configuration, detail, accuracy by each attack and confusion matrix.

Table 2: DDOS evaluation dataset configuration

Attack name	No of Instances
Back	2203
Neptune	107201
Land	21
Smurf	280790
Pod	264
Normal	97277

Table 3: Result table of DDOS detailed accuracy by class

Attack Name	TP Rate	FP Rate	Precision	Recall	F-Measure
back	0.926	0.02	0.908	0.926	0.917
teardrop	0.982	0.02	0.997	0.982	0.989
neptune	0.99	0.01	1	0.99	1
land	0.857	0.03	0.529	0.85	0.655
smurf	0.99	0.01	1	1	1
Normal	0.998	0.001	0.997	0.998	0.997

Table 4: Confusion matrix for DDOS

Classified as	a	b	c	d	e	f	g
a=back	2039	0	0	0	0	0	164
b=teardrop	0	961	0	0	0	0	18
c=nepture	1	0	107187	8	0	0	5
d=land	0	0	3	8	0	0	5
E=smurf	0	0	0	0	28076	6	23
F=pod	0	0	0	0	2	177	85
G=normal	205	3	0	8	18	5	97038

The proposed IDS does not need to have a knowledge of each attack which is practically a difficult task, the system only required knowledge on category of attack into some extends. Therefore, proposed approach allows to combine the power of the Bayesian network learning and also the domain expert knowledge to full extend to detect all types of attacks in network data. Some detection approaches use event timing to construct Bayesian network models where these approaches are not based on actual real time traffic data which can be manipulated easily by an attacker. However, in our research approach we use basic network traffic data to build Bayesian network models.

6. Conclusion

This research proposes a novel distributed IDS to detect and prevent attacks such as denial service, probes, user to root and remote to user attacks. In this work, we implement and evaluate an IDS based on Bayesian network classification modelling technique. The proposed model has an anomaly based IDS system with an adaptive learning process. The proposed IDS system has been evaluated against the KDD DAPRA dataset which

was designed for network IDS evaluation. The research methodology consists of four different Bayesian networks as classification models, where each of these classifier models are interconnected and communicated to predict on incoming network traffic data. Each designed Bayesian network model is capable of detecting a major category of attack such as denial of service (DoS). Experiments and project implementation used WEKA Java API to develop the IDS model and required simulation scenarios. Experiments result shows that proposed IDS system's. The experiments show that the proposed system is effective in detecting attacks in the test dataset and is highly accurate in detecting all major attacks recorded in DARPA dataset. Furthermore, the practical implementation of the proposed IDS system can be utilized to train and detect attacks in live network traffic.

References

Abouzakhar, N. , Gani A., Abuitbel, M. (2010) "Bayesian Learning Networks Approach to Cybercrime Detection", Proceedings of the 2003 PostGraduate Networking Conference, Vol 1, UK

Kunz, C., Tahmasebi, N. , Risse, T. (2011) "Detecting Credential Abuse in the Grid using Bayesian Networks, Grid Computing (GRID)", 2011 12th IEEE/ACM International Conference, Vol 8, Lyon

Khor, K. , Ting C., Amnuaisuk S., (2009) "From Feature Selection to Building of Bayesian Classifiers A Network Intrusion Detection Perspective"", American Journal of Applied Sciences, Malaysia

Tang, K.,Tian M., Wang, Z. (2009) Insider Cyber Threat Situational Awareness Framework using Dynamic Bayesian Networks, Computer Science & Education, 2009. ICCSE '09. 4th International Conference

Howard, G., Bagchi, S., Lebanan, G. (2009) Determining Placement of Intrusion Detectors for a Distributed Application through Bayesian Network Modeling, 11th International Symposium, Volume 5230 , pp 271-290 ,RAID 2008, Cambridge, MA, USA

Debar, M., Dacier, Wespi, A. (2000), A revised taxonomy for intrusion-detection," Annals des Telecommunications, Vol. 1,

Lincoln labotary, (2000.), "DARPA Intrusion Detection Data Sets,[online], http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html.

Jemili, .F, Zaghdoud, M. , Ahmed, M. (2007) "A Framework for an Adaptive Intrusion Detection System using Bayesian Network", IEEE symposium, New Brunswick, NJ

Kendall, .K, (1999) A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems, Massachusetts Institute of Technology

Frigault, M., Lingyu W., (2008) "Measuring Network Security Using Bayesian Network-Based Attack Graphs," Computer Software and Applications, 2008. COMPSAC '08. 32nd Annual IEEE International , vol., no., pp.698,703

Cemerlic, A. , Yang, L. , Kizza, M. (2009) "Network Intrusion Detection Based on Bayesian Networks"

Conrady, S. , Jouffe, L. , (2011) "Introduction to Bayesian Networks Practical and Technical Perspectives"

Wee, Y. , Cheah, W. , Tan, S. (2011)" Causal Discovery and Reasoning for Intrusion Detection using Bayesian Network," International Journal of Machine Learning and Computing (IJMLC), Vol.1(2), 185-192 ISSN: 2010-3700

Korb, K., Nicholson, A. Nicholson, (2004) "Bayesian Artificial Intelligence", [online] , http://www.csse.monash.edu.au/bai/

Wang, Z. , Gombault S. , Guyet, T. (2006)"Towards fast detecting intrusions using key attributes of network traffic," Telecom Bretagne, france, Vol. 1

Kandeeban, S. , Rajesh R.. (2011) "A Mutual Construction for IDS Using GA ," International Journal of Advanced Science, Vol. 29

Mukhopadhyay, I., Chakraborty, M., Chakraborty, S. (2011) "A Comparative Study of Related Technologies of Intrusion Detection and Prevention Systems," Journal of Information Security, Vol. 29

Pr-owl, "Bayesian Netowrk", http://www.pr-owl.org/basics/bn.phpSans Institute, (2011) "Neural Networks to Intrusion Detection," Journal of Information Security

A Near-Miss Management System to Facilitate Forensic Investigation of Software Failures

Madeleine Bihina Bella, Jan Eloff and Martin Olivier
University of Pretoria, Pretoria, South Africa
mbihina@yahoo.fr
jan.eloff@sap.com
martin@mo.co.za

Abstract: The increasing complexity of software applications can lead to operational failures that have disastrous consequences. In order to prevent the recurrence of such failures, a thorough post-mortem investigation is required to identify the root causes involved. This root cause analysis must be based on reliable digital evidence to ensure its objectivity and accuracy. However, current approaches to failure analysis do not promote the collection of digital evidence for causal analysis. A promising alternative is offered by the field of digital forensics. Digital forensics uses proven scientific methods and principles of law to determine the cause of an event based on forensically sound evidence. However, being a reactive process, digital forensics can only be applied after the occurrence of costly failures. This limits its effectiveness, as volatile data that could serve as potential evidence may be destroyed or corrupted after a system crash. A more proactive approach to digital forensics is therefore required. The analysis of near misses is a promising solution to the above issue. Unlike failures, near misses do not result in loss. Instead, they are high-risk situations with the potential for loss or damage, and as such, are often forerunners to serious failures. The detection of near misses therefore provides an opportunity to safely collect relevant failure-related data before the actual failure occurs. Near-miss analysis has been implemented successfully for decades in many engineering disciplines, but it is not yet readily used in the IT industry. The current paper therefore proposes the architecture of a near-miss management system suitable for the software industry by proposing a definition of a near miss from an IT system perspective. The proposed definition is based on the allowed downtime indicated in the Service Level Agreement (SLA), which specifies the system's contractually agreed performance level. The downtime-based definition of near misses is then used to detect and classify near misses based on their risk level.

Keywords: failure, near miss, digital forensics, SLA, downtime

1. Introduction

Software failures have occurred since the beginning of the computer age as is evidenced by the number of highly publicised IT accidents reported in the media. One recent example of a crisis caused by a software bug is the system failure at the Royal Bank of Scotland (RBS), a major bank in the UK, in December 2013. Due to an unspecified technical glitch, the bank's various electronic channels were unavailable and customers were unable to make payments or withdraw cash with their debit cards (Finnegan, 2013). This outage occurred on "Cyber Monday", which is considered one of the busiest online shopping days of the year. This failure is not the first experienced by RBS. In June 2012, another major outage occurred and left millions of customers unable to access their bank accounts for four days due to a failure in a piece of batch-scheduling software. As a result, deposits were not reflected in bank accounts, payrolls were delayed, credit ratings were downgraded and utility bills were not paid (Worstall, 2012).

The examples quoted above show that software failures result in downtime and poor system performance, both of which can have disastrous consequences. Preventing the recurrence of such disasters is therefore crucial. A thorough post-mortem investigation is required to determine the root cause and correct it. To ensure the validity of its results, the investigation must be based on reliable digital evidence such as log files, system settings and database dumps.

Sound evidence of the software failure promotes the objectivity and comprehensiveness of the investigation, which translates into a greater accuracy of the results. Furthermore, reliable evidence is valuable in the event that the software failure leads to a product liability lawsuit. However, current informal approaches to failure analysis do not promote the collection and preservation of digital evidence. Rather than depending on objective evidence analysis, failure analysis methods often rely on the investigator's experience with the system to identify the cause of the problem. Examples of such failure analysis approaches include application performance management tools, transaction management tools, the war room method, as well as troubleshooting (Neebula.com, 2012).

Troubleshooting in particular, which is usually the first response to a system failure, focuses on restoring the system to its operational state as quickly as possible. This allows little time and resources to collect evidence of the failure. Besides, system restoration often requires rebooting, which destroys or tampers with valuable information that could pinpoint the root cause of the problem (Trigg & Doulis, 2008). Both these issues leave the system vulnerable to the recurrence of a similar failure.

The 2012 outage at RBS referred to earlier illustrates the above point. Although the origin of the problem was initially identified through troubleshooting, the patch that was subsequently applied some hours later did not lead to a full system recovery (Du Preez, 2012). This suggests that the troubleshooting approach did not correctly identify the root cause of the problem, and the real source of the glitch was discovered only after a comprehensive investigation by an independent expert had been ordered by the Financial Services Authority (Finnegan, 2013).

In order to ensure that the failure analysis is based on reliable evidence, the investigation must follow a process that favours the collection and analysis of such evidence. The forensic process is well suited to this purpose, as forensic science uses objective evidence to reconstruct past events. The field of digital forensics, as the application of forensic science to digital systems, also follows established procedures meticulously to ensure the accuracy and completeness of digital evidence and to interpret it objectively (Vacca & Rudolph, 2010). We therefore argue that although it is currently limited to criminal events, security incidents and civil matters, digital forensics can serve as an effective alternative for investigating major software failures. Various standards such as the ISO/IEC 27002 information security standard (ISO/IEC, 2007) and the American National Institute of Standards and Technology (NIST) support this view (Scarfone et al., 2008) and recommend the collection of forensic evidence for the investigation of system failures.

However, as beneficial as it is, digital forensics adopts a reactive approach that can only be applied after the occurrence of a failure. This limits the effectiveness of the investigation, since volatile data that could serve as potential evidence may be lost after a system crash. In order to address this limitation of digital forensics, we propose that the forensic investigation be started at an earlier stage, *before* the software failure actually unfolds. This requires the detection of high-risk conditions that can lead to a major failure. These forerunners to failures are known as near misses in the risk analysis literature (Jones et al., 1999).

By definition, a near miss is a hazardous situation where the sequence of events could have caused an accident had it not been interrupted (Jones et al., 1999). This interruption can be caused by chance or by human intervention. A simple example of a near miss in everyday life is the case of a driver crossing a red traffic light at a busy intersection at high speed without causing a collision. As a near miss is very close to an entire accident sequence, it provides a fairly complete set of data about the potentially ensuing accident. Such data can be used as evidence to reconstruct the impending accident and to identify its root cause. The analysis of near misses can therefore offer a proactive way to collect relevant failure-related data before the failure actually occurs. For this reason, we explore the application of near-miss analysis to facilitate the forensic investigation of software failures.

Although not yet used in the software industry, near-miss analysis is an established field successfully used in several engineering fields and in the military and healthcare industries. It is usually performed through so-called near-miss management systems, which involve automated tools designed to record and analyse near misses (Phimister et al., 2003). The current paper therefore presents our work-in-progress for the architecture of a near-miss management system designed for the software industry.

The proposed architecture follows the digital forensic process to collect evidence of the near-miss and determine its root cause. The architecture accommodates design challenges specific to the software industry where the concept of a near miss is still largely unexplored. As illustrated in Figure 1, the architecture therefore lies at the intersection of failure analysis, digital forensics and near-miss analysis.

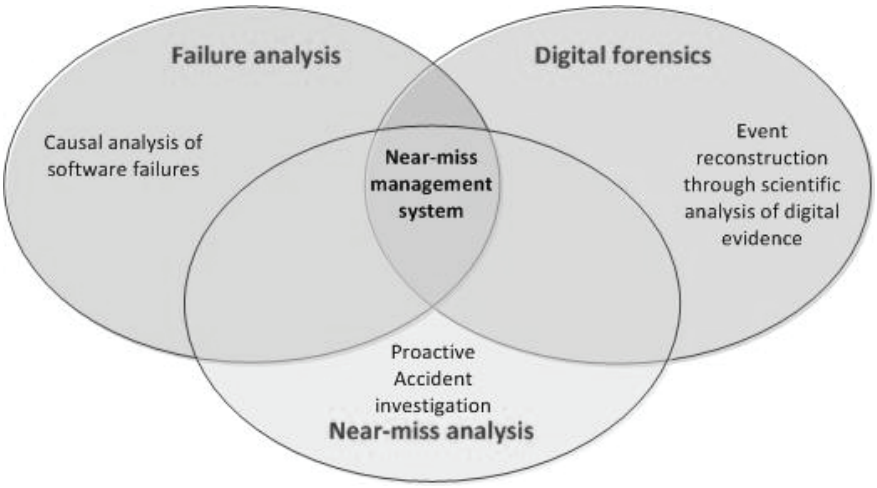

Figure 1: The proposed near-miss management system in relation to the fields of failure analysis, digital forensics and near-miss analysis

The remainder of the paper is organised as follows: Section 2 discusses challenges to near-miss management. Section 3 reviews previous work on near-miss management to address the identified challenges and assesses its suitability for IT systems. Section 4 presents our proposed solutions to the above challenges from an IT system perspective. Finally, an architectural representation of the overall near-miss management system designed to implement the proposed solutions is presented in Section 5.

2. Challenges to near-miss management

Across industries, the successful implementation of a near-miss management system faces two main challenges: the detection of events and conditions that can be classified as near misses, and the high volume of near misses.

2.1 Detection of near misses

Since what exactly constitutes a near-miss event varies between industries and organisations, a clear definition and specification should be given of which type of hazardous events should be reported as near misses. In many industries, near misses are often obtained from observed physical events and conditions. However, such an exercise is particularly challenging in the software industry – in the case of software applications, some near misses might not be visible at all, as no system failure occurs.

2.2 High volume of near misses

Near misses are typically more frequent than failures and can be numerous. In fact, an extensive study of industrial accidents conducted in 1969 already indicated that one actual severe injury can have up to 600 near misses as precursors (Nichol, 2012). (See the popular accident ratio triangle in Figure 2.) A more recent study conducted in 2003 suggests that this number could be even higher (Nichol, 2012).

Figure 2: Bird's accident ratio triangle, adapted from Nichol (2012)

Likewise, a high volume of near misses is expected in the software industry, as reports of various major software failures indicate that these disasters are often preceded by several minor events. However, with limited investigative resources available, the sheer volume of near misses can be virtually unmanageable. An effective data reduction mechanism is therefore required to investigate only those near misses that offer the most complete digital evidence of the potential failure.

Finding solutions to these challenges has been the focus of previous work on near-miss management, which is reviewed in the next section.

3. Previous work on near-miss management

Typically, a near-miss management system focuses on and performs the following three consecutive main processes:

- Identification of near misses
- Selection and prioritisation of near misses for analysis
- Causal analysis of the selected near misses

As mentioned earlier, the first two are particularly challenging, especially for software systems. Moreover, previous work on near-miss identification is very industry-specific and not really applicable to the software industry. We therefore review previous work on the last two processes only so as to identify previous solutions that could be used in the design of the near-miss management system that we propose in this paper.

3.1 Selection and prioritisation of near misses for analysis

Various approaches are used to classify and prioritise near misses. Besides manual screening, risk-based classification and mathematical modelling can be used.

Risk-based classification ranks near misses based on the severity of their potential consequences or on their frequency (Greenwell et al., 2003). This ranking can be performed with a risk decision matrix (Ritwik, 2002) or by analysing historical data of reported near misses to determine trends in the occurrence of certain near-miss events (Phimister et al., 2004).

Mathematical modelling can be used to estimate the conditional probability of an accident given a near miss in order to assess its level of severity. This accident probability is defined as the level of closeness between the near miss and the likely accident. Probabilistic risk analysis (PRA) and Delphi techniques can also be used for the same purpose. PRA consists of estimating the risk of failure of a complex system by breaking it down into its various components and determining potential failure sequences. The Delphi method is a group decision-making tool that can be used to obtain information on the probability of an accident from a panel of experts (Phimister et al., 2004).

Although the above techniques can provide useful results, previous work in near-miss prioritisation is generally specific to the industry concerned and often requires prior knowledge about near misses from historical data. The latter is not available in the software industry, where the concept of near miss is still new. However, one valuable concept that can be applied in the IT industry is the prioritisation of near misses based on the likelihood of a failure. Our suggestion on how to use this approach is presented in Section 4.

3.2 Causal analysis of the selected near misses

Causal analysis of near misses can be performed with investigation techniques borrowed from engineering disciplines – for instance fishbone diagrams, event and causal factor diagrams, event tree analysis, fault tree analysis and failure mode and effects analysis (RealityCharting.com, 2013). The investigation involves answering a series of questions that give insight into the factors that led to the near miss, the possible adverse consequences of the near-miss and the factors that prevented or limited those consequences. The analysis can be conducted by using various tools such as a comparative timeline to organise data and various matrices such as the missed-opportunity matrix and the barrier-analysis matrix (Corcoran, 2004).

Statistical analysis has also been proposed for learning from near misses. Some examples make use of estimation techniques, simulations and regression analysis (Phimister et al., 2004). However, as valuable as

these techniques are, they do not follow forensic principles. Thus they are not suitable for our proposed near-miss management system, which applies the digital forensic methodology to analyse near misses, the same way we suggest using digital forensics to investigate software failures. (Our proposed process for this purpose was described in a previous paper (Bihina Bella et al., 2011).)

As the above review shows, some work is still required to define, classify and prioritise near misses from an IT system perspective. Our vision on how this can be done is explained in the next section.

4. Definition and prioritisation of near misses for IT systems

4.1 Definition of near misses for IT systems

The exact definition of a near miss depends on the industry and organisation involved. However, when referring to engineering products, a common understanding is that near misses are risky conditions or events that can lead to a major failure if not remedied in time (Jones et al., 1999).

With regard to IT systems, a failure is "the inability of a system or component to perform its required functions within specified performance requirements" (IEEE, 1990). Since no system is immune to malfunction, no system vendor can guarantee that the system will work perfectly and continuously at all time. In other words, some periods of downtime are expected. For this reason, the performance requirements of a typical IT system make provision for a downtime "allowance". This allowed downtime can be indicated informally in the system's specifications document, but it is usually formally specified in a contract between the service provider and the receiver of the service (customer). This contract is referred to as a service level agreement (SLA) (Sevcik, 2008).

For instance, the SLA for a website may specify that the site will be operational and available to the customer at least 99.9% of the time in any calendar month. This indicates that the website should not be down for more than 0.1% of the time in a month. For a 30-day month, this corresponds to a downtime limit of 0.03 day or 43 min and 12 s. If the website is down for more than this amount of time in a month, it does not meet the customer's expectation in terms of the SLA. It violates the SLA and is considered to have failed.

Therefore, for the purposes of the research in hand, an event is considered a failure if its resulting downtime exceeds the downtime allowance specified in the SLA. Similarly, an event is considered a near miss if it can lead to exceeding that allowance. We therefore propose the following informal definition of a near miss:

A near miss is an unsafe event or condition that causes a downtime whose duration is close to exceeding the downtime allowance specified in the SLA.

We propose to use the above concept as the basis to formally define a near miss. In order to do so, we need to determine how close the experienced downtime should be from exceeding the allowed downtime to be considered a near miss. We suggest specifying a near-miss threshold for this purpose. This threshold will vary from one organisation to the next, depending on its risk tolerance. For instance, a 95% threshold (95% of the downtime allowed) would correspond to a total monthly downtime of 41 min and 2 s. In this case, a near miss is any monthly downtime of between 41 min and 2 s (the threshold) and 43 min and 12 s (the allowed downtime). This threshold-based definition of a near miss can be mathematically expressed as follows.

$D_{experienced}$ is the experienced downtime
$D_{allowed}$ is the SLA downtime allowance
α is the near-miss threshold in percentage; $\alpha < 1$
$\alpha \times D_{allowed}$ is the near-miss threshold in time value
If $\alpha \times D_{allowed} \leq D_{experienced} \leq D_{allowed}$ then $D_{experienced}$ is a near miss

Figure 3 shows the downtime-based classification of events explained above.

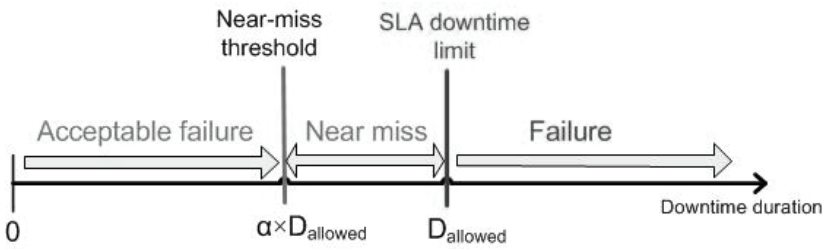

Figure 3: Classification of unsafe events based on their downtime duration

The above formal definition of a near miss enables the detection of near misses among periods of downtime. Once near misses have been identified, they need to be classified and prioritised for data collection and analysis. The next section explores our suggestion on how to prioritise near misses.

4.2 Near-miss prioritisation

We previously defined a near miss as a potential failure, more specifically as an event that can lead to the violation of the SLA. We also defined the violation of the SLA in terms of the system downtime. According to the same logic that we used to measure the severity of a failure based on the downtime experienced, we can assess the severity of a potential failure or near miss based on its expected downtime. In other words, we want to determine for how long the system will be down in the event of an outage caused by this near miss.

To this effect, we need the failure probability of the near miss and the expected recovery time for the outage, known as the MTTR (mean time to repair). The expected downtime is then calculated as the product of the failure probability and the MTTR. The MTTR can be obtained from the system vendor specifications or through historical observations. We are currently exploring formulas from the reliability theory of IT systems (Holenstein, 2003) to develop a suitable failure probability formula.

The system enters a "critical zone" when the expected downtime is greater than the SLA downtime allowance. This can be expressed as the following formula:

$D_{expected}$ is the expected downtime due to a failure
$D_{allowed}$ is the SLA downtime allowance
P is the probability of failure, given the current unsafe situation
MTTR is the expected recovery time following an outage
$D_{expected} = P \times MTTR$
If $D_{expected} > D_{allowed} \rightarrow$ critical zone

Only events in the critical zone are passed on for analysis. These events are prioritised based on the value of their expected downtime. If we manage to prevent a system outage once the system is in the critical zone, then this event is classified as a near miss, otherwise it is a failure. Both cases need to be investigated to identify their root cause and prevent their reoccurrence. In the case of the failure, the root cause analysis is facilitated by the fact that the event data collection occurred before the system failed.

The near miss management system that integrates the above steps is presented in the section below.

5. Architectural representation of near-miss management system

This section describes the architecture proposed for a near-miss management system that implements the steps for detecting and prioritising near misses, as well as the data collection and causal analysis described earlier. It starts with an overview of the overall near-miss management process and is followed by a detailed description of the architecture designed to automate this process.

5.1 The near-miss management process

To detect near misses, the system needs to be monitored with a view to recording and reviewing event logs. It is suggested that events be logged in a central repository such as a Syslog server. By default, Syslog messages are assigned a facility code and a severity value (Gerhards, 2009). The facility code indicates the source of the message (e.g. printer, network) and the severity value indicates its criticality (e.g. warning, error, emergency).

This information can be used for a primary categorisation and ranking of unsafe situations. Only severe events and messages that indicate a status change (e.g. up, down, restart) from critical resources are labelled as potential near misses. Events flagged as critical are then sent to another module for prioritisation. This module calculates the system failure probability and expected downtime based on each event. Afterwards, events identified as being in the critical zone are passed on to another component for data collection. The Simple Network Management Protocol (SNMP) is proposed for this purpose. In our case, the SNMP manager is the data collection module and it requests additional information about the event from the relevant units of the monitored system through their SNMP agents. Corrective steps are subsequently taken to prevent a system failure, if possible. Finally, the collected data is used for causal analysis of the event. Once its root cause has been identified, recommendations are made to correct the system flaw. The architecture of a near-miss management system that is proposed for implementing the above process is described in the next section.

5.2 The near-miss management system architecture

The proposed near-miss management system architecture is shown as a UML component diagram in Figure 4. The architecture consists of five main components or modules listed as follows in their logical sequence:

- The Near-Miss Monitor
- The Near-Miss Classifier
- The Near-Miss Data Collector
- The Failure Prevention
- The Event Investigation

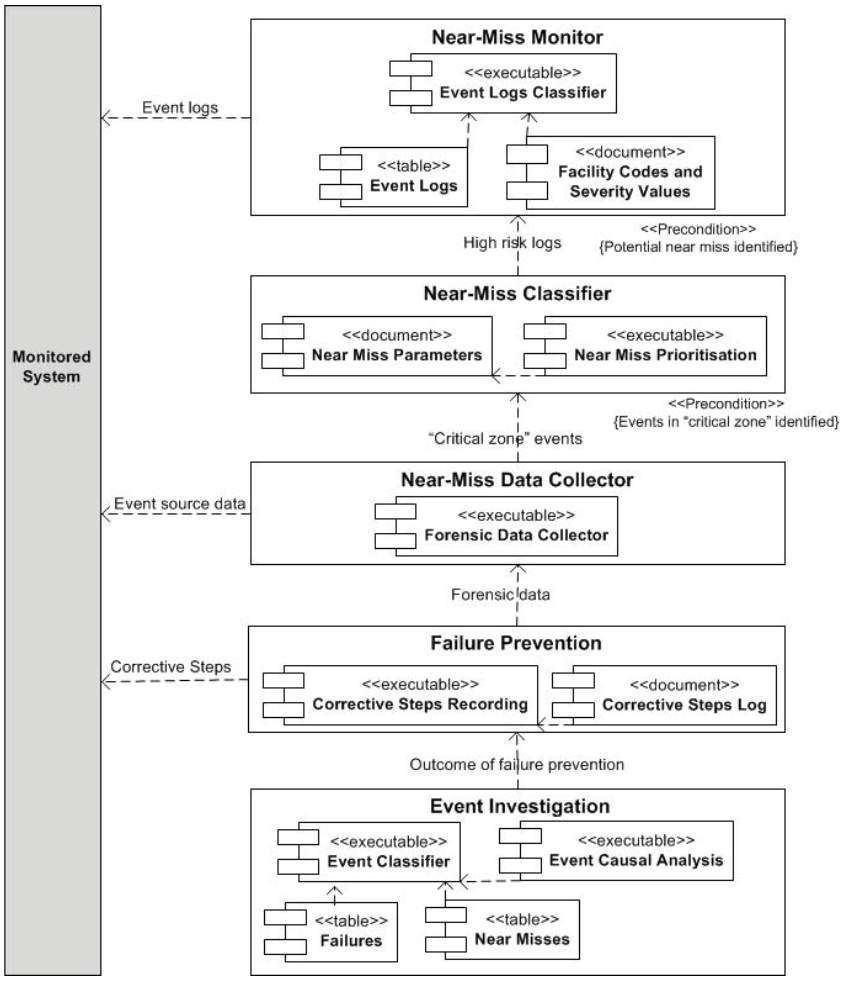

Figure 4: UML component diagram of the near-miss management system

Some components consist of several sub-components that all have a type: a document file, an executable file or a database table. Dashed arrows indicate a dependency from the source component to the target

component. For instance, the Near-Miss Classifier requires high-risk event logs from the Near-Miss Monitor. Some dependencies are subject to conditions, for example an expected downtime must occur in the critical zone to activate a data request from the Near-Miss Data Collector.

Each of the main modules processes the event logs from the monitored system. The five main components of the system are used to perform a multi-stage filtering process that progressively discards "irrelevant" events and only retains near misses with the highest risk factor. The key component of the architecture is the module used to prioritise near misses, known as the Near-Miss Classifier. It uses the equation established in Section 4.2 to classify near misses based on their expected downtime. A description of the main components of the architecture follows.

5.2.1 The near-miss monitor

The Near-Miss Monitor monitors the units of the system to identify potential near misses. This module is implemented as a Syslog server that receives logs from every unit. Using predefined criteria (message content, source and risk level), the Syslog server classifies the logs based on their facility codes and severity values. Events classified as potential near misses are then sent to the Near-Miss Classifier.

5.2.2 The near-miss classifier

The Near-Miss Classifier calculates the failure probability and expected downtime, and prioritises events accordingly. Logs of events identified as being in the "critical zone" are sent to the Near-Miss Data Collector and an alarm is raised to notify the system administrator.

5.2.3 The near-miss data collector

The Near-Miss Data Collector module is implemented as an SNMP Manager that requests data from the units in the critical zone. Such data may include the source identifier (e.g. IP address), running processes, system settings and error messages. This data is then stored in the Event Data table before being transferred to the Failure Prevention module.

5.2.4 The failure prevention

In the Failure Prevention module, the system administrator uses the collected data to identify and implement appropriate corrective steps in an attempt to prevent or at least mitigate the impact of a system failure. This might include ending some active but unused processes or deleting some stored but unnecessary data to free up memory. The administrator records the steps implemented in a log file for future reference. He then sends the outcome of the recovery attempt (successful or unsuccessful) to the Event Investigation module.

5.2.5 The event investigation

Based on the outcome of the failure prevention process, the Event Investigation module classifies events as either near misses or failures and stores the event details in the appropriate table for future reference. The administrator then conducts a root cause analysis of the event based on the data stored. Afterwards, recommendations for improvement are made and implemented - either immediately or at a later scheduled time. The recommendations are stored along with the event details in the relevant table. These steps allow for the creation of an event history that can be looked up in the event of a similar event occurring in the future.

The above architecture enables the proactive collection of data related to the potential failure to be used for root cause analysis. Besides, it enables the improvement of the software correctness once the cause of the outage has been established and recommendations for improvement have been implemented. This prevents the recurrence of a similar failure in the future. An additional benefit of the architecture is that it enables the prevention of a failure if appropriate corrective actions are executed timely.

6. In conclusion

This paper examines the application of a near-miss management system to facilitate the forensic investigation of severe software failures. The paper established a definition of near misses for IT systems, and developed a method to detect, classify and prioritise near misses so as to address the issue caused by their high volume. The automation of this process, which was presented through the architecture of a near-miss management

system, is work-in-progress and research is underway to refine it. Future work is aimed at developing a mathematical formula to calculate the likelihood of a failure due to a near miss. This would allow for the prioritisation of high-risk near misses for detailed causal analysis.

References

Bihina Bella, M.A., Olivier, M.S. and Eloff, J.H.P. (2011). *Proposing a digital operational forensic investigation process*, WDFIA, London, UK.

Corcoran. W.R. (2004). "Defining and analyzing precursors", *Accident precursor analysis and management: Reducing technological risk through diligence*, Phimister, J.R., Bier V. and Kunreuther, H., Washington D.C, USA: The National Academies Press, pp 79-84.

Du Preez, D. (2012). "RBS failure to fix IT systems after month is 'unprecedented' say analysts", Computerworld UK, [online] http://www.computerworlduk.com/in-depth/infrastructure/3369901/rbs-failure-to-fix-it-systems-after-month-is-unprecedented-say-analysts/?intcmp=rel_articles;infrstrctr;link_4

Finnegan, M. (2013). "RBS apologises as customers hit by another IT outage", Computerworld UK, [online] http://www.computerworlduk.com/news/it-business/3491865/rbs-apologises-as-customers-hit-by-another-it-outage/

Gerhards, R. (2009). "The Syslog Protocol", RFC 5424, [online] http://tools.ietf.org/html/rfc5424

Greenwell, W.S., Knight, J.C. and Strunk, E.A. (2003). "Risk-based classification of incidents" *in Workshop on the Investigation and Reporting of Incidents and Accidents*, Department of Computer Science, University of Virginia, [online] http://shemesh.larc.nasa.gov/iria03/p03-greenwell.pdf

Holenstein, B., Highleyman, B. and Holenstein, P.J. (2003). *Breaking the availability barrier: Survivable systems for enterprise computing*, Vol. 1, Bloomington, USA: Authorhouse, pp 27-28.

IEEE. (1990). *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries*.

ISO/IEC. (2007). "Information technology - Security techniques - Code of practice for information security management", International Standard ISO/IEC 27002, Switzerland, Geneva.

Jones, S., Kirchsteiger, C. and Bjerke, W. (1999). "The importance of near miss reporting to further improve safety performance", *Journal of Loss Prevention in the Process Industries*, Vol. 12, pp 59-67.

Neebula.com. (2012). "Success Factors for Root-Cause Analysis", [online] e-Book, http://www.neebula.com

Nichol, K. (2012). "The safety triangle explained", [online] http://crsp-safety101.blogspot.com/2012/07/the-safety-triangle-explained.html

Phimister, J.R., Bier, V.M. and Kunreuther, H.C. (2004). *Accident precursor analysis and management: Reducing technological risk through diligence***,** National Academies Press, [online] http://www.nap.edu/catalog/11061.html

Phimister, J.R., Oktem, U., Kleindorfer, P.R. and Kunreuther, H. (2003). "Near-miss incident management in the chemical process industry", *Risk Analysis*, Vol. 23, No. 3, pp 445-459.

RealityCharting.com. (2013). "Various RCA Methods and Tools in Use Today", [online] http://www.realitycharting.com/methodology/conventional-wisdom/rca-methods-compared

Ritwik, U. (2002) "Risk-based approach to near miss". *Hydrocarbon processing,* pp 93-96.

Scarfone, K., Grance, T. and Masone, K. (2008). "Computer security incident handling guide", *NIST Special Publication 800-61*, Revision 1, National Institute of Standards and Technology, Gaithersburg, USA, [online] http://csrc.nist.gov/publications/nistpubs/800-61-rev1/SP800-61rev1.pdf

Sevcik, P. (2008). "Service Level Agreements for Business-Critical Applications", NetForecast Report 5091, [online] http://www.netforecast.com/wp-content/uploads/2012/06/NFR5091SLAsforBusiness-CriticalApplications.pdf

Trigg, J. and Doulis, J. (2008). "Troubleshooting: what can go wrong and how to fix it", *Practical Guide to Clinical Computing – Systems: Design, Operations, and Infrastructure*, Chapter 7, pp 105-128, 2008, Elsevier Inc, London.

Vacca, J.R. and Rudolph, K. (2010). *System Forensics, Investigation and Response*, Chapter 1, pp 2-16. Sudbury, Mass.: Jones & Bartlett Learning.

Worstall, T. (2012) "RBS/NatWest Computer Failure: Fully Explained", [online] http://www.forbes.com/sites/timworstall/2012/06/25/rbsnatwest-computer-failure-fully-explained/

Requirements for Preparing the Cloud to Become Ready for Digital Forensic Investigation

Moses Dlamini[1], Hein Venter[1], Jan Eloff[1, 2] and Mariki Eloff[3]
[1]ICSA Research Group, Department of Computer Science, University of Pretoria, Pretoria, South Africa
[2]SAP Research P&I Mobile Empowerment Africa, Pretoria, South Africa
[3]Institute for Corporate Citizenship, University of South Africa, Pretoria, South Africa
mdlamini@cs.up.ac.za
hventer@cs.up.ac.za
jan.eloff@sap.com
eloffmm@unisa.ac.za

Abstract: Some research work claims that the adoption rates for cloud computing have not scaled as anticipated. This is mainly due to security concerns. Hence, most research efforts have been directed at the cloud security challenges. In the meantime, researchers have rarely focused on an equally important issue: that of digital forensics investigation in the cloud. It is on this premise that this paper makes an attempt to understand the challenges and complexities of conducting an effective digital forensics investigation in the cloud. The contribution is to provide a better understanding of the implications that cloud computing poses for digital forensics investigators. This paper proposes digital forensic readiness as a solution to help prepare the cloud for an effective investigation. The paper goes further to explain and motivate why the authors view digital forensic readiness as the most appropriate solution. It also discusses the major factors that are influencing the move towards digital forensic readiness for cloud computing. As a work-in-progress, this paper ends by proposing necessary requirements that must be considered in order to make the cloud become ready for an effective investigation. It also provides a glimpse to the proposed three-tier digital forensic readiness model for the cloud which is grounded on these fundamental requirements.

Keywords: cloud computing, digital forensics, digital forensic investigation, digital forensic readiness

1. Introduction

Cloud computing is arguably one of the most recent research frontiers across the IT space that is fast moving towards maturity. However, security concerns still remain a formidable challenge for most organisations intending to move their critical data to the cloud (Hoffman, 2013). On this premise, significant research efforts are still being directed at security challenges of the cloud. The research community have rarely focused on digital forensics investigation on the cloud. It is vital to protect and secure cloud systems from the inevitable security threats. However, it is equally important to be able to trace and track the sources of threats in order to identify the malicious culprits and bring them to book. Digital forensics investigation fills that gap. However, quite often it comes into play after an incident has already occurred taking a reactive approach.

Organisations cannot afford to follow the same reactive approach to digital forensic investigations on the cloud. This is mainly due to the transitory and dynamic setup of cloud services. This setup makes it hard for digital forensic investigators to gather credible evidence. However, the same setup also enables attackers to commit crimes and quickly destroy any traces of evidence. Attackers could simply migrate to another cloud service provider where they could do the same and leave without a trace (Feibig et al., 2013; Martini and Choo, 2012). Moreover, the shared cloud environment makes it possible for clients' data to co-reside on a single physical infrastructure. Such co-residence causes isolation or retrieval of one client's resources involved in an investigation to adversely affect others who may not be involved.

Based on the above and other challenges, conducting an effective digital forensic investigation on the cloud has remained a big challenge. Martini and Choo (2012) argue that there is a lack of recent understanding of the challenges that are brought about by cloud computing. Hence, this paper intends to provide a better understanding of the implications of cloud computing on the field of digital forensics. It acknowledges that the rise of cloud computing brings about a digital forensic dilemma. In order to solve this dilemma, this paper proposes a digital forensic readiness approach which is grounded on a number of specific requirements meant to solve the digital forensic crisis on the cloud.

The remainder of the paper is structured as follows: section 2 discusses some background and related work. Section 3 discusses the challenges of conducting a digital forensic investigation in cloud computing environments. Section 4 introduces digital forensic readiness and the requirements thereof on the cloud. Section 5 concludes and provides potential future work.

2. Background and related work

This section provides the theoretical background on which this work is grounded. It defines cloud computing and digital forensics and gives a brief discussion to set the scene. This section ends by reflecting exactly on why the rise of cloud computing is seen as the beginning of a digital forensic crisis.

2.1 Cloud computing definition

The most widely used definition of cloud computing is found in the NIST publication, which defines it as "a model for enabling convenient, elastic and on-demand access to a shared pool of configurable computing resources and services that can be rapidly provisioned and released with minimal management effort" (Mell and Grance, 2011). Substantial work has already been discussed within cloud computing. However, the research community has seldom discussed the issue of digital forensic investigation on the cloud.

2.2 Digital forensics

Digital forensics is a disciple that came about as a result of the ubiquity of digital devices and their increasing use in criminal or malicious activities (Reilly et al., 2011). Some define digital forensics as a scientifically-proven methodology for investigating digital devices which may be suspected to have been involved in criminal or malicious activities for potential evidence that could be used in a court of law in order to help prosecute the perpetrators (Ngobeni et al., 2012). This definition tends to lean more towards a criminal investigation. A comprehensive definition found in several other sources defines digital forensics as "the use of scientifically derived and proven methods towards the identification, collection, preservation, validation, documentation, analysis, interpretation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events suspected to be of a criminal or malicious nature or assisting to anticipate unauthorized actions" (Agarwal et al., 2011; Harrington, 2012; ISO/IEC WD 27043.2, 2014). This definition encompasses both the criminal and corporate investigation aspects. Hence, most of our discussion from hereon would consider both the criminal and corporate investigation. The next section discusses the effects of cloud computing on digital forensics.

2.3 Related work

Conducting an effective digital forensics investigation in order to help prosecute cybercriminals in the cloud is still quite an open challenge. Research on this topic is still insufficient and has not scaled to the expected levels (Grispos et al., 2013; Grispos et al., 2011; Gupta, 2011). There is a need to extend and improve the existing tools and technologies to cater for the new cloud environment. Hence, some researchers Birk (2011), Hare-Brown and Douglas (2011), Ruan et al. (2011), Reilly et al, (2011), Taylor et al., (2010), Sibiya, Venter and Fogwill (2012) have already made attempts to identify some of the digital forensic challenges that arise in the cloud computing environment. The challenges of conducting a digital forensic investigation in a cloud computing environment as gleaned from the related work is summarized by the authors as depicted in table 1.

Below is a legend on how to read table 1.

- ✓ denotes a challenge
- ✘ denotes not a challenge

Of note is that some of the identified challenges are not necessarily specific to a cloud environment, but are also relevant for the traditional non-cloud environment in general. For example, the lack of consistent Internet laws has always been a challenge in both traditional dead and live forensics. It is still a challenge on cloud forensics. Casey (2012) argues that some of the issues that are a cause of concern in cloud environments have already been dealt with in the past. Martini and Choo (2012) asserts that even though the cloud might have introduced new challenges for digital forensics, the existing standards and key principles of digital forensics should be maintained.

Table 1: The challenges of conducting an effective digital forensics investigation on the cloud

Challenge			Dead Forensics	Live Forensics	Cloud Forensics
Large number of potential source of evidence	Massive data collection		×	×	✓
	Search and seizure		×	✓	✓
	Selective capturing/acquisition (proportionality rule)		✓	✓	✓
	Complexity and timescale		✓	✓	✓
	Insufficient storage capacity		✓	✓	×
Virtualization	VM snapshot acquisition		×	✓	✓
	No checksum for image snapshots		×	✓	✓
	No hardware write-block		×	✓	✓
	Non repeatable actions	Lack of integrity	×	✓	✓
		Lack of system re-construction	×	✓	✓
	No physical access		✓	✓	✓
Volatile and Live Data	Lost evidence on shutdown		×	✓	✓
	No bit-by-bit copy		×	✓	✓
	Deleted data overwritten		×	✓	✓
	Obtaining routing information		✓	✓	✓
Multi-tenancy	Co-residence of clients' data		×	×	✓
	Logical segregation controls		×	×	✓
Encrypted data sets			✓	✓	×
Contractual Obligations	Implementation of SLA		×	×	✓
	Internal audits		×	×	✓
	Responsibility to supply forensically sound audit logs		×	×	✓
	Lack of well written contracts and liability issues thereof		×	×	✓
Control of Evidence	Collection and acquisition	Replicated data	×	×	✓
		Remote acquisition	×	×	✓
		Un-cooperating service provider	×	×	✓
	Preservation	Transfer of evidence	✓	✓	✓
		Audit logs for clients	×	×	✓
		Chain of custody	×	×	✓
	Evidence validation		×	✓	✓
	Established method for evidence acquisition		×	✓	✓
	Evidence verification		×	✓	✓
Presentation	Technical jargon		✓	✓	✓
	Evidence weight and admissibility		×	✓	✓
Multi-jurisdiction	No cross cutting legislation		×	×	✓
	Search warrants		×	×	✓
	Legal vs illegal		×	✓	✓
	Lack of consistent Internet laws		✓	✓	✓

Hence, this paper acknowledges that conducting a digital forensic investigation on the cloud environment comes with new challenges along with those already encountered in physical forensics. The next section elaborates on some of the challenges.

3. The rise of cloud computing brings a digital forensic crisis

Cloud computing makes a great business model with all its benefits. However, it also makes a completely unworkable model for digital forensic investigators (Fred Cohen & Associates, 2009). For example, Martini and Choo (2012) argue that the increasing use of cloud services presents both opportunities for criminal and a host of challenges for digital forensic investigators. Cloud computing has made it easy for criminals to exploit storage service providers like Dropbox to store incriminating evidence (Casey, 2012). However, this makes it difficult for digital forensic investigators to seize such data for forensic purposes. For example, some cloud service providers distribute and store client data in multiple jurisdictions. In such cases, obtaining search

warrants and getting the necessary consent from multi-jurisdictional law enforcers in urgent and time-sensitive situations may prove to be not feasible.

The implications for acquiring and preserving digital evidence cannot be ignored (Mason and George, 2011). Some researchers even argue that digital forensics faces one of its greatest challenges with the rise of cloud computing (Sang, 2013; Martini and Choo, 2012; Reilly et al., 2011; Ruan et al., 2011; Yadav, 2011; Zimmerman and Glavach, 2011). Harrington (2012) argues that technological trends such as cloud computing and social media are pushing legal and ethical boundaries to the limits and necessitating novel approaches to digital forensic investigations. For example, an investigator who goes beyond the formal process of acquiring digital evidence to befriend a suspect under false pretence on social networking media in order to gain access to the suspect's private information could easily lead to an *invasion of privacy*, *intrusion upon seclusion* or *unreasonable warrantless search* or other *tort liability* (Harrington, 2012).

Therefore, this paper argues that the rise of cloud computing brings along a digital forensic crisis. New developments in cloud computing are driving the change of approach in digital forensic investigations. For example, standardization efforts have already begun to try and harmonise digital forensic investigation processes. This paper is based on a draft ISO standard titled "ISO/IEC 27043 – Information Technology – Security Techniques – Incident Investigation Principles and Processes" (ISO/IEC WD 27043.2, 2014) which is about to be approved by the time of writing this paper.

The next section proposes digital forensic readiness as one way to prepare the cloud environment for an effective digital forensic investigation.

4. Digital forensic readiness

Digital forensics has been traditionally viewed as a reactive approach. There has been insufficient research directed at proactive digital forensics. However, with the rise of critical IT infrastructures and cloud computing systems, there is a considerable amount of research that now focuses on a proactive approach of collecting and preserving digital evidence prior to a potential investigation. This is mainly due to the fact that reactive approaches are insufficient and not scalable in the cloud environment (Chung et al., 2012).

This is a call for cloud-bound organisations to proactively collect potential digital evidence and preserve it in a legally and forensically sound manner. Section 4.1 outlines the main drivers of doing so. Section 4.2 provides a formalized list of requirements that must be considered to help prepare the cloud to become forensically ready for a digital investigation.

4.1 Main drivers

This section discusses some of the forces that are driving most organisations towards considering a digital forensic readiness capability. The main drivers include, but are not limited to, corporate governance and legal requirements, policies and guidelines, costly business disruptions, the duty to preserve digital evidence, and strict courts' obligations to evidence admissibility. These are discussed briefly in the next subsections.

4.1.1 Corporate governance and legal requirements

Cloud-bound organisations could use a digital forensic readiness capability as part of an all-encompassing incident response procedure. This could help to demonstrate good corporate governance and compliance with legal and regulatory mandates. For example, according to the King 3 report on corporate governance, the effective utilization of digital forensic tools can enable cloud organisations to prove their due diligence with respect to good governance (Grobler et al., 2010b).

Furthermore, the ISO 27001/2 standard has a provision that make it essential for organisations to identify and gather potential digital evidence that is complete, admissible and concrete prior to a litigation or dispute to determine the root cause of an incident and make means to timely prosecute the perpetrators (Grobler et al., 2010a). The PCI DSS (Payment Card Industry Data Security Standard) also has an obligation for financial organizations to enable digital forensic processes to help provide for timely forensic investigation in the event of a compromise to any of its cloud-hosted merchants or service providers (Mouhataropoulos et al., 2011).

4.1.2 Digital forensic readiness – a policy matter

Digital forensic readiness has now become a policy matter. For example, some organisations in the United Kingdom (UK) are already implementing digital forensic readiness policies (Irwin, 2012; Mouhtapoulos et al., 2011, Rowlingson, 2004). This came about after a legal mandate was announced for all organizations that are dealing directly or indirectly with the UK government to reasonably anticipate and respond in a forensically-ready manner to any potential incident that might lead to a dispute or litigation. The aim of such a policy is to provide a systematic, standardized and legal basis for the acceptance and admissibility of potential digital evidence that may be required in a formal dispute or legal litigation process. Several researchers (Cobb, 2011; Danielsson and Tjostheim, 2004; Rowlingson, 2004; Grobler and Louwrens, 2007; Tan, 2001) claim that having such a policy in place will maximise organisations' potential to gather credible digital evidence that could be admissible in court whilst minimising the cost of conducting a digital forensic investigation.

4.1.3 Disruption to business

A digital forensic readiness capability should also help organisations to avoid business disruptions during an investigation. Cobb (2011) asserts that the objectives of a digital forensic readiness policy are;

- to maximise the usefulness of legally and ethically acquired admissible digital evidence
- to minimize any costs of an investigation and disruptions to business operations.

A well-structured digital forensic readiness capability presents the potential to significantly reduce the cost and time of an investigation. On the other hand it could also increase the prospects of a quick and successful legal or dispute outcome with minimal disruptions to business activities. For example, in a cloud setting, when all the potential digital evidence is kept at a secure remote site there is absolutely no need for the investigators to interrupt the cloud provider's business operation activities. Investigations could be carried out with minimal business disruption and without raising unnecessary suspicion to the clients being investigated.

4.1.4 Duty to proactively gather and preserve potential digital evidence

A digital forensic readiness capability should enable cloud clients and service providers to take proactive measures towards acquisition; preservation and storage of potential digital evidence in anticipation of a potential litigation and/or disputes that may adversely affect and disrupt business operations. The aim is to proactively gather, preserve and store credible and admissible digital evidence to be ready in case of litigation or dispute in response to "*the duty to preserve*" as obligated by a number of courts of law (Cross and Kuwahara, 2010).

Taylor (2012) argues that several courts require litigants to capture, preserve and produce relevant and potential digital evidence well in advance. Failure to do so could result in heavy sanctions, tort liability and even default judgement against litigants.

4.1.5 Inadmissible digital evidence

There is a need to involve a legal expert to help determine the exact scope of evidence to be gathered and preserved. The legal expert would also determine the legality of the digital forensic tools to be used and outline the necessary steps to gather evidence. All these to ensure that evidence remain admissible in court as per the much discussed Daubert's criteria of determining the reliability and admissibility of scientific evidence (Majmudar, 1993; Mcleod, 2000; Orofino, 1996). This criteria state that scientific and/or digital evidence (1) must be grounded on empirically testable theory or technique; (2) the theory must have been deeply scrutinized and peer-reviewed; (3) its potential error rate should be clearly stated; and (4) it must be based on generally accepted scientific principles (Majmudar, 1993). Otherwise, the collected evidence might be easily considered inappropriate or inadmissible in a court of law; or worse still it might be considered a misconduct or unlawful invasion of privacy by the investigating authority as evident in the *United States vs Maynard* case (Harrington, 2012).

4.2 Digital forensic readiness requirements for the cloud

This section outlines a list of basic requirements to be considered for a successful implementation of a digital forensic readiness capability on the cloud environment. The list of basic requirements is gleaned from the

related work. The requirements provide a sound point of departure on how to prepare the cloud to become forensically ready for digital forensic investigations.

4.2.1 Requirement 1: Minimize the time or cost of acquiring digital evidence on the cloud

Significant time lags during evidence acquisition could easily risk potential digital evidence being corrupted or lost before it could be captured for analysis. Moreover, this problem could be exacerbated by providers' short log retention policies. Hence, for the proposed digital forensic readiness model to be effective, it should be able to significantly minimize the time and/or cost it takes digital forensic investigators to legally acquire potential digital evidence and conduct an investigation on the cloud. A digital forensic readiness model should take a proactive approach towards cost containment ensuring that potential digital evidence is acquired in as minimal time and budget as possible, regardless of the scope of the case.

4.2.2 Requirement 2: Minimize business disruption

Taking cognisance of the fact that a cloud computing environment consist of multiple tenants who are likely to be sharing the cloud infrastructure, it becomes vital that the proposed digital forensic readiness model should ensure that an investigation can be seamlessly carried out without any major disruption or interference to other business activities of the cloud service provider. It should be noted that most clients leave a lot of traces on their access devices when using cloud services. So, if at all possible, investigators should first focus on the client access devices and only move to the cloud service provider when more concrete evidence is required or when the available evidence is not conclusive. This should be done according to the 'proportionality doctrine' as explained in (Harrington, , 2012).

4.2.3 Requirement 3: Demonstrate due diligence and compliance with corporate governance, legal and regulatory mandates

The penalty for non-compliance to corporate governance, legal and regulatory mandates goes beyond simple monetary fines to criminal and/or civil litigation. For example, the PCI-DSS states that a level one merchant that fails to comply with the standard's requirements could be fined an amount within the range of $500,000 - $1,000,000. The PCI-DSS also states that inability to pass an audit and show any mitigation could potentially lead to the merchant losing its right to accept credit card charges (Centrify Corporation, 2012). Hence, an effective digital forensic readiness model should strive to demonstrate due diligence and compliance to corporate governance, legal and regulatory requirements. Therefore, it is necessary that legislation and other mandates should be considered to ensure that digital evidence is admissible in court.

4.2.4 Requirement 4: Securely and selectively gather, store and preserve legally admissible digital evidence

The cloud environment is mainly characterised by its elasticity and scalability to cope with just any amount of data. Nonetheless, it is unreasonable to expect cloud-based organisations to try and preserve all potential digital evidence they have in their possession. The cost of doing so would be unreasonable and not feasible at all. Hence, organisations should balance this requirement with the "proportionality doctrine" in terms of the cost, time and technological viability of doing so to that of only selectively capturing what is potentially relevant and necessary. It is imperative for such organisations to consider the scope of potential digital evidence acquisition; the cost and time involved; the burden and disruption of business activities; and the defensibility of the process of acquisition. The acquired potential evidence should be stored in a secure repository with tightly-controlled access, preferably on a remote site. The requirements for digital forensic readiness in the cloud can be summarized as follows.

In another paper, the authors are proposing a three-tier digital forensic readiness process model for the cloud which is grounded on the above requirements as shown below.

The three tiers of the proposed model consist of a planning, implementation and assessment tiers. Each tier consists of a number of processes aimed at achieving digital forensic readiness on the cloud. The details of the model are contained in another article and are not part of this paper. The next section concludes this paper and provides directions for future research work.

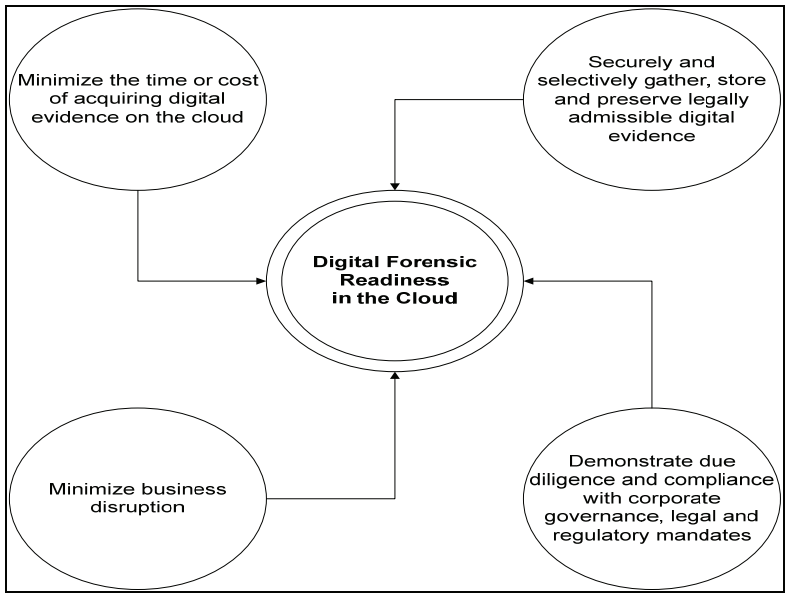
Minimize the time or cost of acquiring digital evidence on the cloud
Securely and selectively gather, store and preserve legally admissible digital evidence
Digital Forensic Readiness in the Cloud
Minimize business disruption
Demonstrate due diligence and compliance with corporate governance, legal and regulatory mandates

Figure 1: Digital forensic readiness requirements for the cloud

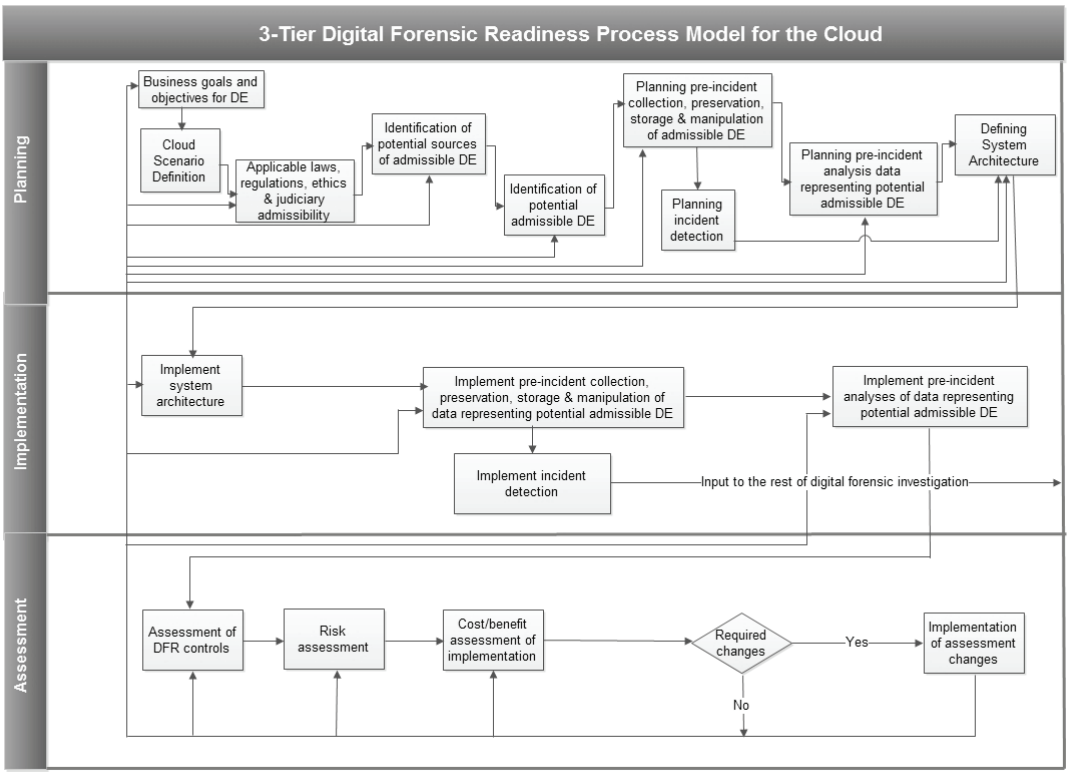
3-Tier Digital Forensic Readiness Process Model for the Cloud
Planning
Business goals and objectives for DE
Cloud Scenario Definition
Applicable laws, regulations, ethics & judiciary admissibility
Identification of potential sources of admissible DE
Identification of potential admissible DE
Planning pre-incident collection, preservation, storage & manipulation of admissible DE
Planning incident detection
Planning pre-incident analysis data representing potential admissible DE
Defining System Architecture
Implementation
Implement system architecture
Implement pre-incident collection, preservation, storage & manipulation of data representing potential admissible DE
Implement pre-incident analyses of data representing potential admissible DE
Implement incident detection
Input to the rest of digital forensic investigation
Assessment
Assessment of DFR controls
Risk assessment
Cost/benefit assessment of implementation
Required changes
Yes
No
Implementation of assessment changes

Figure 2: A 3-tier digital forensic readiness model for the cloud

5. Conclusion

Cloud computing presents an innovative way to deliver IT resources over the Internet. However, it also comes with some challenges for digital forensic investigators. This paper provides an understanding of the challenges of digital forensics in the cloud environment and the implications thereof. The authors deduced that a traditional reactive digital forensic approach is neither effective nor efficient to gather credible digital evidence on the cloud and proposed a proactive approach. Hence, this paper has outlined a new set of requirements that must be considered in order to proactively gather digital evidence from the cloud. These requirements are

being considered for the implementation of a 3-tier digital forensic readiness process model as part of future work.

References

Agarwal, A., Gupta, M., Gupta, S. and Gupta, S.C. (2011). Systematic Digital Forensic Investigation Model, *International Journal of Computer Science and Security (IJCSS)*, vol.5, no.1, 2011, pp.:118 – 131.

Birk, D. (2011). Technical Challenges of Forensic Investigations in Cloud Computing Environments, *Workshop on Cryptography and Security in Clouds, IBM Forum,* 15 – 16 March 2011, Zurich, Switzerland.

Casey, E.(2012). Cloud Computing and digital forensics, *Journal of Digital Investigation,* vol. 9, no. 2012, pp.: 69 -70.

Centrify Corporation (2012). Using Microsoft Active Directory to Address Payment Card Industry (PCI) Data Security Standard Requirements in heterogeneous Environments, Whitepaper, *2012 Centrify Corporation,* available at: http://resources.idgenterprise.com/original/AST-0076426_centrify_wp011_active_directory_and_pci.pdf, accessed: 09 December 2013.

Chung, H., Park, J., Lee, S. and Kang, C. (2012). Digital forensic investigation of cloud storage services, *Journal of Digital Investigation, Elsevier Ltd.* Vol. 9, no.2012, pp.: 81 – 95.

Cobb, M. (2011). Digital forensic investigation procedure: Form a computer forensic policy, *TechTarget, ComputerWeekly.com,* available at: [http://www.computerweekly.com/tip/Digital-forensic-investigation-procedure-Form-a-computer-forensics-policy], accessed: 19 November 2013.

Cross, D.D. and Kuwahara, E. (2010). E-Discovery and Cloud Computing: Control of ESI in the Cloud, *EDDE Journal,* vol.1, no.2, Spring 2010, pp: 2 – 12.

Danielsson, J. and Tjøstheim, I. (2004). The Need for a Structure Approach to Digital Forensic Readiness: Digital Forensic Readiness and E-commerce, IADIS International Conference e-Commerce, 2004, available at: http://el.trc.gov.om:4000/htmlroot/ENGG/tcolon/e_references/Consolidated/Computer%20Science/Journals/The%20need%20for%20a%20structured%20approach%20to%20digital%20forensic%20readiness.pdf, accessed 08 October 2013.

Dlamini, M.T., Venter, H.S., Eloff, J.H.P. and M.M. Eloff (2011). Security of Cloud Computing: Seeing Through the Fog, *SATNAC 2011,* 4-7 September 2011, East London, South Africa

Feibig, S., Siebanhaar, M., Gottron, C. and Steinmetz, R. (2013). VM Live Migration using a Hybrid External Approach, 4th International Conference on Cloud Computing and Service Science (CLOSER, 2013), Aochin, Germany, 8 -10 May 2013.

Fred Cohen & Associates (2009). Analyst Report and Newsletter, *Fred Cohen & Associates,* available at: http://all.net/Analyst/2009-05b.pdf, accessed 27 November 2013.

Grispos, G., Glisson, W.B. and Storer, T. (2013). Using Smartphones as a Proxy for Forensic Evidence contained in Cloud Storage Services, 46th Hawaii International Conference on System Science, IEEE Computer Society, pp.:4908 - 4919

Grispos, G., Storer, T. and Glisson, W. B. (2011).Calm before the storm: The emerging challenges of cloud computing in digital forensics. August 2011. [.pdf], available at: http://www.dcs.gla.ac.uk/~tws/papers/grispos11calm-rev2425.pdf, accessed: 09 September 2013.

Grobler, T. and Louwrens, B.(2007), in IFIP International Federation for Information Processing, Volume 232, New Approaches for Security, Privacy and Trust in Complex Environments, eds. Venter, H., Eloff, M., Labuschagne, L., Eloff, J.,von Solms, R., (Boston: Springer), pp. 13-24.

Grobler, C.P., Louwrens, C.P.and von Solms (2010a), A framework to guide the implementation of Proactive Digital Forensics in organizations, *2010 International Conference on Availability, Reliability and Security, Appears in IEEE Computer Society,* pp. 677 – 682.

Grobler, C.P., Louwrens, C.P. and von Solms (2010b), A multi-component view of digital Forensics, the Proceedings of the 2010 International Conference on Availability, Reliability and Security, appears in IEEE Computer Society, pp. 647 – 652.

Gupta, U. (2011). Forensics in the Cloud: 5 Hot Skills; Today's Virtual Environment Puts New Demands on Investigators, *Bank Information Security Articles*, available at: http://www.bankinfosecurity.com/articles.php?art_id=4091&opg=1, accessed: 04 October 2013.

Hare-Brown, N and Douglas, J. (2011). Digital investigations in the Cloud, *QCC Information Security Ltd, Digital Forensics Laboratories Whitepaper*, available at: www.qccis.com/whitepapers, accessed: 22 August 2011.

Harrington, S.L. (2012). Collaborating with a Digital Forensics Expert: Ultimate Tag-team or Disastrous Duo? *William Mitchell Law Review*, vol.38, no.1, January 2012, pp: 353 – 396.

Hoffman, S. (2013). Survey: Security Still a Cloud Stumbling Block, available online:channelnomics.com/2013/05/24/survey-security-cloud-stumbling-block/, accessed 25 March 2014.

Irwin, R. (March 2012). NHS Lancashire Forensic Readiness Policy, available at: http://www.centrallancashire.nhs.uk/Library/Documents/policies/information-governance-policies/NHS%20Lancashire%20Forensic%20Readiness%20Policy%20March%202013.pdf, accessed: 31 October 2013.

ISO/IEC WD 27043.2 (2012). Information Technology – Security Techniques – Investigation principles and processes Standard (Draft), *Secretariat of ISO/IEC JTC 27/WG 4,* DIN German Institute for Standardization, Berlin.

Majmudar, K.B. (1993). Daubert v. Merrell Dow: A Flexible Approach to the Admissibility of Novel Scientific Evidence, *Harvard Journal of Law & Technology*, Vol. 7, No. 1, Fall 1993, pp. 187 – 205.

Martini, B. and Choo, K.R. (2012). An integrated conceptual digital forensic framework for cloud computing*, Journal of Digital Investigation*, Vol.9, No. 2012, pp.:71 – 80.

Mason, S. and George, E. (2011). Digital evidence and' cloud' computing, *Computer Law & Security Review, Elservier Ltd*, Vol. 27, No. 2011, pp.:524 -528.

Mell P., and Grance, T. (2011). The NIST Definition of Cloud Computing (Draft): Recommendations of the National Institute of Standards and Technology, Special Publication 800-145 (Draft), January 2011, available at: http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf, accessed: 10 September 2012.

Mcleod, A.K. (2000). Is Frye Dying or Is Daubert Doomed? Determining the Standard of Admissibility of Scientific Evidence in Alabama Courts, *Alabama Law Review,* Vol. 51, No. 2, pp. 883 – 905.

Mouhataropoulos, A., Grobler, M. And Li, C.T. (2011). Digital Forensic Readiness: An insight into Governmental and Academic Initiatives, *2011 European Intelligence and Security Informatics Conference (EISIC 2011),IEEE Computer Society,* Athens, Greece, 12 – 14 September 2011, pp.191 - 196.

Ngobeni, S., Venter, H.S. and Burke, I. (2012). The Modelling of a Digital Forensic Readiness Approach for Wireless Local Area Networks, *Journal of Universal Computer Science(J.UCS)*, vol.18, no.12, 2012, pp: 1721 – 1740.

Orofino, S. (1996). Daubert v. Merrell Dow Pharmaceuticals, Inc.: The Battle Over Admissibility Standards for Scientific Evidence in Court, *Journal Undergraduate Science, History of Science,* Vol. 3, Summer 1996, pp. 109 -111.

Reilly, D., Wren, C. And Berry, T. (2011). Cloud Computing: Pros and Cons for Computer Forensic Investigations, *International Journal Multimedia and Image Processing (IJMIP),* vol.1, no.1, March 2011, pp: 26-34.

Rowlingson, R. (2004). A Ten Step Process for Forensic Readiness, *International Journal of Digital Evidence,* vol.2, no.3, Winter 2004, pp.:1- 28.

Ruan, K., Carthy, J., Kechadi, T. and Crosbie, M. (2011). Cloud Forensics: An Overview, available at: http://cloudforensicsresearch.org/publication/Cloud_Forensics_An_Overview_7th_IFIP.pdf, accessed: 10 September 2013.

Sang, T. (2013). A Log-based Approach to Make Digital Forensics Easier on Cloud Computing. 3rd International Conference on Intelligent System Design and Engineering Applications. Hong Kong, 16 -18 January 2013, pp. 91 – 94.

Sibiya, G., Venter, H.S. and Fogwill, T. (2012). Digital Forensic Framework for a Cloud Environment, *IST-Africa 2012 Conference Proceedings, Paul Cunningham and Miriam Cunningham (Eds), IIMC International Information Management Corporation, 2012, ISBN: 978-1-905824-34-2.*

Tan, J. (2001). Forensic readiness, *Cambridge, MA: Stake; 2001.*

Taylor, K. (2012). Cloud Consideration: E-Discovery, *K&L Gates,* available at: http://www.legalcloudcentral.com/2012/10/articles/in-the-courts/cloud-considerations-ediscovery/#more, accessed: 27 November 2013.

Taylor, M., Haggerty, J., Gresty, D. and Hegarty, R. (2010). Digital Evidence in Cloud Computing Systems, *Computer Law & Security Review Journal, Elsevier Ltd,* Vol. 26, pp: 304 – 308.

Yadav, S. (2011). Analysis of Digital Forensic and Investigation, *VSRD International Journal of Computer Science & Information Technology (VSRD-IJCSIT),* vol.1, no.3, pp.: 171 – 178.

Zimmerman, S. and Glavach, D. (2011). Cyber Forensics in the Cloud, *IA newsletter for Information Assurance Technology Professionals*, Winter 2011, vol. 14, no. 1, pp: 4 – 7.

A Generic Framework for Enhancing the Quality Digital Evidence Reports

Nickson Karie[1, 2] and Hein Venter[1]
[1]Department of Computer Science, University of Pretoria, Pretoria, South Africa
[2]Department of Computer Science, Kabarak University, Kabarak, Kenya
menza06@hotmail.com
hventer@cs.up.ac.za

Abstract: In recent times, the admissibility of potential digital evidence in any court of law is coming under increased scrutiny. This is aggravated by the fact that, the complexity of digital evidence is also increasing on a daily basis. Thus, convincing the court that the evidence presented is worthy of inclusion into any criminal process, the digital forensic experts require extensive technical knowledge and skills, including methodologies and specifications typically designed for producing quality digital evidence reports. This also implies that, the techniques, knowledge and skills used by the digital forensic experts during the preparation of digital evidence reports, should have the ability to convince the court on the validity, reliability and weight of the evidence captured during the investigation process. The methodologies used should also be able to assist the law enforcement agencies determine, with less effort, the admissibility of the digital evidence being reported. As of the time of this study, however, there exists a lack of a standardised or harmonised framework that have specifically been designed to help in the process of preparing quality digital evidence reports after an investigation has been conducted. This has, therefore, resulted in disparities on how digital evidence reports are prepared and presented in any court of law or civil proceedings. To address this disparity problem, a unified process equipped guidelines that meets some specified minimum requirements on how digital evidence reports should be prepared would be worthwhile. This paper, therefore, presents a step-by-step framework in an attempt to propose high-level guidelines that can be adapted to enhance the process of preparing digital evidence reports. The use of such a framework, for example, can be a stepping stone towards a harmonised process on how to prepare digital evidence reports for use in court or legal proceedings. Besides, such a framework can also assist the law enforcement agencies, for example, to determine, with less effort, the reliability, validity, weight and admissibility of any digital evidence included in the final reports.

Keywords: digital forensics, methodologies, digital evidence reports, framework, high-level guidelines, harmonised process

1. Introduction

The use of different types of digital evidence in court systems has increased in the past few decades. The complexity of digital evidence is also increasing on a daily basis. In the case of a digital forensic investigation process, for example, the different types of digital evidence that can be captured includes among others information on desktop computers, laptops, server, mobile phones, audio files, video recordings, emails, web logs, digital images, short message services and voice conversations. Convincing the court that any of such digital evidence captured is worthy of inclusion into the criminal process, the digital forensic experts require extensive technical knowledge and skills, including methodologies and specifications typically designed to help them prepare quality digital evidence reports for use in court.

Besides, as digital crimes continue to increase and overwhelm digital forensic investigators and crime laboratories, law enforcement agencies also need a sure way to determine the validity, weight and admissibility of any digital evidence presented in court. For this reason, standardised methodologies and specifications need to be developed in digital forensics to assist the law enforcement agencies determine, with less effort, the validity, weight and admissibility of any digital evidence reported.

As of the time of this study, however, there exists a lack of standardised or harmonised digital forensic processes that have specifically been designed to help in preparing quality digital evidence reports for use in court. This has, therefore, resulted in disparities on how digital evidence reports are prepared and presented to different stakeholders, more specifically, after a digital forensic investigation process has been conducted.

The proposed framework in this paper is, thus, an attempt towards proposing high-level guidelines that can help enhance the process of preparing quality digital evidence reports for use in any legal proceedings. Furthermore, the requirement for such methodologies and specifications in digital forensics is exceptionally important - both for the advancement of the field as well as for the effective use of tools, upon which the science of digital forensics and use in evaluation by courts depend (Cohen, 2011). Such methodologies will also

assist law enforcement agencies, for example, in differentiating between experts' own opinions from the actual depiction of the digital evidence (Karie and Venter 2013). The framework can also be useful to the digital forensic experts, for example, in drawing inferences from the digital evidence itself. Finally, this paper is meant to spark discussions and further research on an international agreed framework to enhance the process of preparing quality digital evidence reports for use in legal proceedings.

The remaining part of this paper is organised as follows: section 2 presents background concepts while section 3 handles previous related work. A detailed explanation of the proposed framework is handled in section 4 followed by a critical evaluation of the framework in section 5. Finally, conclusion and future work is given in section 6.

2. Background

According to Casey (2004), digital evidence is any probative information stored or transmitted in digital form that can be used at a trial in any court of law. Various digital forensic tools and techniques are normally used by investigators to capture data from computer systems and other digital devices in an attempt to identify potential digital evidence that can provide conclusive descriptions of the activities that took place. Nonetheless, before accepting any of the captured digital evidence, the court has to determine if such evidence is relevant, authentic, hearsay and whether a copy is acceptable or the original is required (Casey, 2004).

As stated by Sherman (2006), digital forensic experts can discover significant and damning evidence that can potentially convict suspects and prosecute them. However, no matter how momentous the evidence or how skilled the investigator has been at recovering it, if the digital evidence reports are not presented in a coherent and understandable way to the court, the case may be lost. For this reason, in the case of a digital forensic investigation process, the presentation of digital evidence reports become one critical phase to any digital forensic investigator. This is because; it is in this phase where information about the results and findings of the investigation process has to be reported to the relevant stakeholders (Sherman, 2006).

Nevertheless, in presenting the final digital evidence reports investigators are more often than not challenged by the fact that there is currently a lack of standardised guidelines that one must adhere to when preparing digital evidence reports. Poorly prepared digital evidence reports can, for example, make it hard to convince the court that the defendant is guilty of the crime he or she is accused of (Karie and Venter, 2013).

Despite the advances in digital forensics, legal professionals and researchers are yet to resolve the challenges associated with digital evidence reports presented in court. Existing methodologies and specifications have not addressed this problem fully with the result being that, individuals might be denied justice due to lack of guidelines for preparing quality digital evidence reports for use in legal proceedings (Karie and Venter, 2013). Thus, methodologies need to be developed in digital forensics to address the challenges and disparities associated with digital evidence reports prepared for use in any court of law or legal proceedings. In the next section, we examine existing related work in digital forensics.

3. Related work

There exists various research works from different digital forensic researchers which have made valuable contributions towards the proposed framework in this paper. In this section, therefore, a summary of some of the most prominent efforts in previous research work is provided.

To begin with, Hamda et al. (2011) states that, due to the lack of standards in reporting digital evidence items, investigators often face difficulties in efficiently presenting investigation findings. Thus, they propose in their paper a standard for digital evidence to be used in reports that are generated using digital forensic tools. Based on the investigation findings, the standard for digital evidence reports can include items such as data about the case, the evidence source, evidence item, and the chain of custody. However, in the current paper we focus on presenting high-level guidelines that can be used to enhance the process of preparing quality digital evidence reports for use in court.

In another effort by Boddington et al (2008) they argue that digital evidence is now common in legal cases. However, the understanding of the legal fraternity as to how far conventional ideas of evidence can be

extended into the digital domain lags behind. There arises a need, therefore, for a practical 'roadmap' that can guide the legal practitioner in identifying potential digital evidence relevant to support a particular case and in assessing its weight. Their paper goes further and describes a process by which the validation of relevant potential digital evidence required for legal argument can be facilitated, by an interrogative approach that ensures the chain of reasoning is sustained. Their paper, though, does not discuss any specific steps to be followed when preparing final digital evidence reports for use in legal proceedings.

Another effort by Karie and Venter (2013) presents a framework in an attempt to propose high-level guidelines for enhancing the presentation of potential digital evidence in any legal proceedings. Such a framework can be useful to the digital forensic experts, for example, in structuring investigation findings as well as in identifying relevant patterns of events to be incorporated during the presentation of digital evidence in court. The framework can also assist law enforcement agencies; for example, determine with less effort, the validity, weight and admissibility of any potential digital evidence presented in court. However, the framework focusses more on the presentation phase and does not discuss any specific guidelines to be used when preparing digital evidence reports after a digital investigation process has been conducted, which is the focus of this paper.

There also exist other related works on issues related to digital evidence and evidence reporting but neither those nor the cited references in this paper have presented a step-by-step framework with guidelines to enhance the process of preparing quality digital evidence reports for use in any court of law or legal proceedings in the way that is introduced in this paper. However, we acknowledge the fact that the previous research works have offered useful insights toward the development of the framework in this paper. In the section that follows, we explain in more detail the proposed framework.

4. The proposed framework

In this section of the paper, the authors present a detailed explanation of the proposed framework. Figure 1 shows the structure of the framework.

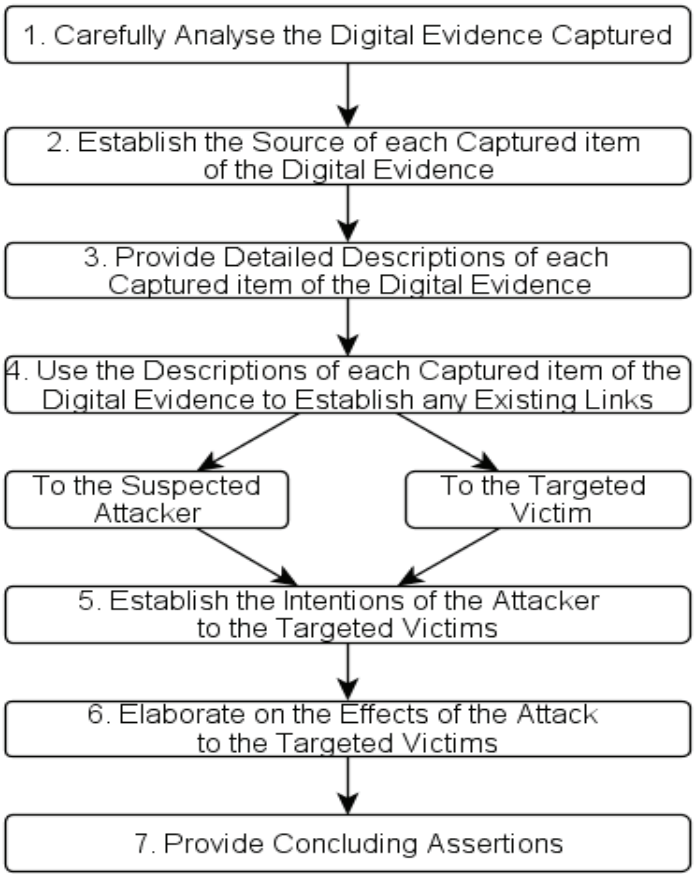
1. Carefully Analyse the Digital Evidence Captured
2. Establish the Source of each Captured item of the Digital Evidence
3. Provide Detailed Descriptions of each Captured item of the Digital Evidence
4. Use the Descriptions of each Captured item of the Digital Evidence to Establish any Existing Links
To the Suspected Attacker
To the Targeted Victim
5. Establish the Intentions of the Attacker to the Targeted Victims
6. Elaborate on the Effects of the Attack to the Targeted Victims
7. Provide Concluding Assertions

Figure 1: A framework for enhancing the quality of digital evidence reports

The framework consists of seven proposed steps which can be adapted to enhance the quality of digital evidence reports. The steps are arranged from top to bottom where the first step is to carefully analyse the digital evidence captured. This is followed by establishing the source of each captured item of the digital evidence in the second step. The third step provides detailed descriptions of each captured item of the digital evidence. The fourth step uses the descriptions of each captured item of the digital evidence to establish any existing links to the suspected attacker or targeted victim. Step five establishes the intentions of the attacker to the targeted victims based on the descriptions of the captured digital evidence. Elaborating on the effects of the attack to the targeted victims is presented in step six of the framework. Finally, concluding assertions are supplied in step seven.

Note that all the steps proposed in the framework shown in Figure 1 demand the use of scientifically-proven methods. Such methods are beyond the scope of this paper; however, when used in digital forensics, they must be based on empirical and measurable evidence subject to specific scientific principles. In the sub-sections to follow, the steps 1 to 7 as shown in Figure 1 are explained in more detail.

4.1 Carefully analyse the digital evidence captured

In writing the final digital evidence report, investigators must first carefully analyse the different types of digital evidence captured during the investigation process. The digital evidence may include among others types; audio files, video recordings, emails, web logs, digital images, short message services and voice conversations. Because of the sheer magnitude of the different types of digital evidence that can be captured, investigators should only concentrate on analysing evidence that is relevant to the case at hand. This helps the investigators in managing the evidence information to be included in the final report.

Besides, before using any of the captured digital evidence in the final report to determine the truth of an issue, the investigator must be sure that such digital evidence has been captured and carefully analysed. Having analysed the different types of evidence before compiling the final report can be a confidence booster to the digital forensic expert, especially on the inferences drawn from such digital evidence (Karie and Venter, 2013). Evidence analysis makes it easier to establish the source of each captured item of digital evidence, which forms step two of this framework.

Note that the process of capturing any digital evidence demands the use of scientifically-proven methods. As mentioned earlier, such methods are not discussed this paper, but, when used in digital forensics, they must be based on empirical and measurable standards subject to specific scientific principles.

4.2 Establish the source of each captured item of the digital evidence

With the advances in digital technology, the sources of digital evidence have also grown exponentially. It is, therefore, important that investigators establish reliable sources for each of the different types of digital evidence captured during the investigation process before compiling the final reports. Failure to establish reliable sources of the captured digital evidence, for example, can make it hard for such evidence to be considered for inclusion in the final report.

Having multiple reliable sources of the same digital evidence, for example, can be used to establish the continuity of the offense as well as any existing links to related digital evidence. This is useful especially, when more than one digital system has been used in committing the offense. As stated by Casey, (2012), the more corroborating evidence that investigators can capture, the more certainty they can have in their final conclusions. Thus, having unreliable evidence sources can potentially be more damaging than having no sources.

All relevant and reliable evidence sources should, hence, be established and considered in the final digital evidence report. This will particularly allow multiple sources of evidence to be analysed as well as ensure the suitability of including such digital evidence in the report

4.3 Provide detailed descriptions of each captured digital evidence

In the case of a digital forensic investigation process, many of the different types of evidence captured are entirely in digital form. Providing detailed descriptions of each of the captured evidence can be worthwhile. Such descriptions could further be used in supporting the prosecution of different types of digital crimes.

However, in describing the digital evidence captured in the final report, the investigators may also be required to explain the process used to analyse the digital evidence itself. This is important especially when trying to establish the authenticity of the evidence. To be sure that each item of the digital evidence incorporated in the final report is explained, investigators can create a checklist that identifies all the digital evidence items considered in the report.

4.4 Use the descriptions of the captured digital evidence to establish any existing links

Establishing any existing links between captured digital evidence can reveal the relationships between such digital evidence to the suspected attacker or targeted victim as shown in step four of Figure 1. Based on the crime committed it is possible that some of the digital evidence captured may have little or no links to either the suspected attacker or the targeted victim. This, however, can be based on the weight, validity, reliability and the inferences drawn from the potential digital evidence itself as described in step three of Figure 1.

4.5 Establish the Intentions of the attacker to the targeted victims

Discovering the intention behind the digital attack can lead to discovering the different ways in which the attack was done. Different evidence captured can, thus, be used to prove the attacker's intention, for example, compromising the confidentiality, integrity, or availability of the victim's data among others. Trying to figure out what the attacker's intention is can also help in establishing the effects of the attack. For this reason, based on the captured and analysed digital evidence, the investigator should, if possible, state clearly in the final report the intentions of the attacker to the targeted victims and the effects there after. Step six of Figure 1 discussed next is, therefore, proposed in the framework to help elaborate on the effects of the attack to the targeted victims.

4.6 Elaborate on the effects of the attack to the targeted victims

Many of the perpetrators of digital crimes usually targets particular victims or services hosted by the victims. Different types of attacks may also have different impacts or effects on the targeted victims. It is, therefore, essential that digital forensic investigators elaborate, in the final evidence reports, the impacts or effects of the attack to the targeted victim. Such explanation can assist the court in drawing reasonable inference that can as well help in reasoning and identifying digital evidence relevant to support or refute a particular criminal case in court.

In the case of supporting an existing case in court, the digital forensic expert might as well be required to further show in their final report that the support is as a result of an existing relationship (link) between one, two or more of the potential digital evidence artifacts captured during the investigation process. Although when elaborating on the effects of the attack to the targeted victim, the digital forensic experts should also indicate whether including all the details of the findings that support or refute a particular criminal case in the final report are absolutely necessary to the law enforcement requirements (Karie and Venter, 2013).

4.7 Provide concluding assertions

Finally, the seventh and the last step of the proposed framework provide concluding assertions. This means that the final report should include any important details from each of the steps as shown in Figure 1. The reference protocols followed and the methods used, to seize, document, collect, preserve, recover, reconstruct, organize and search for any key digital evidence may also need to be elaborated (Casey, 2004) as concluding assertions. Step seven may also include supporting or refuting some of the captured digital evidence in the final report. However, the support or refusal of any digital evidence should be based on the estimated weight, validity, reliability and the inferences made from such digital evidence during the analysis process. The next section, presents a critical evaluation of the framework proposed in this paper.

5. Critical evaluation of the proposed framework

The proposed framework in this paper is a new contribution in the digital forensics domain. The scope of the framework is defined by the steps and guidelines as shown in Figure 1. The main steps as depicted in the framework include:

- Carefully Analyse the Digital Evidence Captured
- Establish the Source of each Captured item of the Digital Evidence
- Provide Detailed Descriptions of each Captured item of the Digital Evidence
- Use the Descriptions of each Captured item of the Digital Evidence to Establish any Existing Links
- Establish the Intentions of the Attacker to the Targeted Victims
- Elaborate on the Effects of the Attack to the Targeted Victims
- Provide Concluding Assertions

The specific details of the individual steps as identified in the framework have further been explained in this paper. However, note that the steps as identified in Figure 1 are meant to facilitate this study and primarily focus on enhancing the process of preparing quality digital evidence reports for use in any court of law or legal proceedings. Such proposed steps or guidelines are by no means the final guaranteed steps to digital evidence reports admissibility in court. Nevertheless, organising the framework into steps was necessary to simplify the understanding of the framework.

Some of the problems that led to the introduction of this new framework in this paper include:

- Lack of a common legal basis in preparing digital evidence reports: The lack of a common legal basis for preparing quality reports can easily render digital evidence reports inadmissible in court. For this reason, having a common legal basis for preparing digital evidence reports can help digital forensic investigators focus on the key issues to be included in the final evidence reports to be presented at trials.
- Lack of a standardised or harmonised procedure for preparing digital evidence reports: Standardising procedures for preparing digital evidence reports after a digital investigation process has been conducted can reduce the time taken by investigators to prepare court admissible reports. This is backed up by the fact that investigators will only concentrate on evidence that relates to the case at hand while forgetting the rest. The framework in this paper, therefore, is a step towards a harmonised process for preparing quality digital evidence reports.
- Reliability/trustworthiness of digital evidence reports: It is not always obvious that the reliability of digital evidence will be acceptable in court; however, the use of a harmonised framework, for example, can be a positive step towards producing acceptable level of trustworthiness and reliability in digital evidence reports. The reliability can also be questionable if the data included in the final reports are not accurate and trustworthy. Therefore, the use of common and/or standardised frameworks to prepare digital evidence reports can help raise the level of reliability and/or trustworthiness of digital evidence reports.

The proposed Framework as demonstrated in this paper can, thus, be used in the digital forensics domain, for example, to help investigators in preparing quality digital evidence reports as well as in identifying relevant digital evidence to be incorporated in the final digital evidence reports. The framework can also be helpful to law enforcement agencies and other stakeholders, for example, in reasoning and identifying digital evidence relevant to support or refute a particular criminal case presented in court.

For the case of digital evidence admissibility in legal proceedings, the steps as identified in the framework shown in Figure 1 can also be useful, for example, in evaluating the validity, reliability and weight of the digital evidence included in the final report. The framework can, further, be used for training investigators, especially on the art of digital evidence report writing for use in any court of law or civil proceedings.

Academic institutions should also find the framework in this paper constructive, especially when training students on how to write quality digital evidence reports for use in legal proceedings. Such a framework can assist in developing curriculums and education materials for different programs of study within the field of digital forensics. Such programs will, for example, ensure that institutions produce well-enabled digital forensic specialists capable of writing high quality digital evidence reports.

Developers of digital forensics tools can also use the proposed framework to develop automated digital evidence reporting tools. This also implies that developers might find the framework in this paper useful, especially when considering the development of new digital forensic tools and techniques for addressing the disparities experienced during the preparation of digital evidence reports for use in legal proceedings.

6. Conclusion

The problem addressed in this paper was that of the lack of a standardised or harmonised process specifically designed to help investigators in preparing quality digital evidence reports for use in any court of law or legal proceedings. This is backed up by the fact that there is currently a lack of standardised guidelines that investigators should adhere to when preparing final digital forensic evidence reports. This scenarion has brought about disparities in the process followed when preparing digital evidence reports.

A framework is then proposed in this paper in an attempt to provide high level guidelines for enhancing the process of preparing quality digital evidence reports for use in legal proceedings. The requirement of such a framework in digital forensics is exceptionally important to digital forensic investigators, especially during the preparation of final digital evidence reports. Such a framework can also assist the law enforcement agencies, for example, to determine, with less effort, the validity, weight and admissibility of digital evidence incorporated in the report.

Moreover, the framework can also help law enforcement agencies, for example, to differentiate between experts' own opinions from what the digital evidence really portrays. The ability to differentiate opinions from the real digital evidence presented can as well assist the court in evaluating opinions that substantially outweighs prejudicial effect. Again, the framework can be useful in determining the most relevant and appropriate digital evidence to be included in digital evidence reports.

Finally, the authors believe that by using such a framework, quality reports of digital evidence in any court or legal proceedings can be attained. Other future relevant undertakings in the digital forensics domain might as well benefit from applying such a framework as the one proposed in this paper. However, more research needs to be conducted in order to improve on the proposed framework in this paper. The framework should also spark further discussion on the development of new methodologies and techniques to enhance the process of preparing quality digital evidence reports for use in any court of law or legal proceedings.

References

Boddington, R., Hobbs, V., and Mann, G., (2008). Validating digital evidence for legal argument. In the Proceedings of the 6th Australian Digital Forensics Conference, Edith Cowan University, Perth Western Australia

Casey, E. (2012). Error, Uncertainty, and Loss in Digital Evidence. International Journal of Digital Evidence. Vol. 1, No. 2.

Casey, E.(2004). Digital Evidence and Computer Crime, Second Edition. Published by Elsevier. ISBN 0-12-163104-4

Cohen, F., (2011). Digital Forensic Evidence Examination, 3RD Edition. Published by Fred Cohen & Associates ISBN # 1-878109-46-4

Hamda, B., Mariam H., and Ibrahim, B., (2011). Defining a Standard for Reporting Digital Evidence Items in Computer Forensic Tools. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Vol. 53, pp78-95

Karie, N.M. and Venter, H.S., (2013). Towards a Framework for Enhancing Potential Digital Evidence Presentation. In the Proceedings of the 13th Annual Information Security for South Africa Conference, (ISSA-2013). Johannesburg, South Africa. Published online by IEEE Xplore®.

Sherman, S., (2006). A digital forensic practitioner's guide to giving evidence in a court of law. In the Proceedings of the 4th Australian Digital Forensics Conference, Edith Cowan University, Perth Western Australia.

A Framework to Address Challenges Encountered When Designing a Cyber-Range

Brendan Lawless, Jason Flood and Anthony Keane
Institute of Technology Blanchardstown, Dublin, Ireland
brendanlawless2000@gmail.com
jasoneflood@gmail.com
anthony.keane@itb.ie

Abstract: The national infrastructure of modern countries have demonstrated a high dependence on vulnerable automated computer based technologies. At the core of the problem is the convergence of networked systems to TCP/IP everywhere along with the proliferation of smart devices entering the home has increased the cyber attack surface and radically altered the risk profile of end users, organizations and governments alike. It is of critical importance that the individuals responsible for ensuring the security of Cloud, SCADA, smart meters and other intrinsically important pieces of a nation's infrastructure are aware of all threats, vulnerabilities and exploits in the system they protect. This paper proposes that a modular Cyber-Range Framework is a valuable training asset that can be leveraged to help protect critical infrastructure. A Cyber-Range is a vehicle used to train in offensive and defensive Information Operations and Information Warfare. Skills transfer of security principles through gamification facilitates the calculation of metrics that describe the skill level of the player taking part in Cyber-Range activities. Gap analysis based on security player accomplishment will promote the creation of new targets inside the Cyber-Range that focus on addressing an individuals or organizations shortfalls. New challenges may be required due to a recognizable skills gaps or technical advancements in the marketplace. It is proposed that a readily adaptable framework for the design of Cyber-Ranges can be built to cater for current and future cyber defensive needs. Key to this ideology is developing the Cyber-Range Framework in a modular way. Using this framework new challenges can be considered as simple plugin modules that are easily turned on, off and fine-tuned. This methodology will allow for growth and can be focused to suit an individual, organizational or national requirement. Although this is still a work in progress, it has already been used as the backbone of the highly successful Global OWASP 2013 CTF. Design challenges, implementation challenges and the future of the proposed framework are discussed in this paper. Comparisons are drawn between the current in place solutions and our modular Cyber-Range Framework. The benefits and challenges of both approaches are discussed.

Keywords: cyber-range, modular cyber-range, security, gamification

1. Introduction

Typical Cyber-Range Goals are (Cory Janssen, 2014):

- Replicate large scale, complex and diverse networks and its users for future and current systems and operations
- Enable a realistic testing facility for Internet and Global Information Grid (GIG) research
- Enable the development and deployment of state-of-the-art cyber testing capabilities
- Facilitate the scientific use of cyber testing methods
- Provide a virtual environment for the quantitative, qualitative and realistic assessment of potentially ground-breaking cyber technologies for research and development

The approach adopted for the Cyber-Range framework outlined in this document is a modular co-operative system. There are several benefits to this approach.

- Cost: The cost of the equipment can be shared amongst all the partners of the project.
- Expertise: Being a modular design means that the Cyber-Range can grow with the available skills of the contributors.
- Maintenance: Workload around ensuring uptime can be spread between partners.
- Expansion: The modular Cyber-Range format will allow for levels to be hot swapped in and out to suit the audience.

The current construction is a highly challenging Cyber-Range that can be accessed remotely or on-site. The range itself demonstrates a growing number of attack vectors from SCADA controlled electrical systems, server exploits, application exploits to mal-configured encryption technologies. The competent learner can engage in

a number of malicious cyber activities and exploit the range and refine their skills. Using gamification as an educational vehicle learners are encouraged to progress through the framework to capture more points.

Although this is still a work in progress early results are very promising. Using the Global OWASP CTF as a measure of interest and skillset. We believe this framework has the capacity to change how security principles are taught and cyber skills are transferred. We plan to launch the framework publicly at the Irish IRISS 2014 conference.

2. What is a cyber-range?

A Cyber-Range is a real world implementation of combining multiple tiers of the technical stack that enables the playing out of a cyber attack scenarios from both the attacker and defender perspective. Traditional Capture the Flag challenges provide interest for the technical community according to (Department of Defence, 2012) which may be too narrow a field to bring about real change. A Cyber-Range bridges to divide between the real and cyber world. This visual and tactile environment increases awareness of both the impacts and risks of a cyber attack to wider less technical audience. Exploits in the range are designed to showcase complex attack vectors such as those used against national grid or state asset to individual members of the virtual Cyber-Range population. Cyber-Ranges are generally maintained by definitive leaders in the information technology field due to their skills pool requirement.

3. Current approach considerations to cyber-range implementations

The first consideration when implementing a Cyber-Range is the physical scale that the Cyber-Range ultimately achieves for its real world representation. The next logical step is to find a suitable housing location for the Cyber-Range, based on the chosen scale. Unlike most cyber constructs, a range is a bridge between the digital and the physical. The location will need to be accessible with access to key infrastructure components.

If the Cyber-Range will be utilized to trial new software or hardware or 3rd parties will be invited to participate in a Cyber-Range event then locational access policies may need to be amended to allow 3rd party physical access to the Cyber-Range site. Once the location is determined, the actual components that the Cyber-Range structure is to be constructed from must be determined.

Once the physical size and raw materials necessary to build a Cyber-Range to the desired technical scope has been determined, it is now time to look at the back bone systems. Virtualisation of servers is recommended but not yet a standard. Considerations need to be given regarding the various technical components, services and server interactions. Future proofing API design at the start will strongly limit re-working of code bases. Rapid prototyping as a software development model although tempting in the context of a Cyber-Range should be avoided.

4. Management of cyber-range activities

Effective management of a Cyber-Range is a difficult but essential requirement as discussed by (Michael Rosenstein et all, 2012) which describes a “Range Controller” that enables a level of quality to be enacted over the Cyber-Range domain. Guaranteeing a level of operational availability is a key component in the evolution of a Cyber-Range and must not be considered a bolt on component that can be added at the end. Reporting systems must proactively alert Cyber-Range management of the events going on inside the range as this information will enable required steps to be taken to protect the ranges operation.

Building an accessible, stable and clean environment that is available on demand throughout the lifetime of a ranges deployment is important. It is easy for player frustration caused by poor user experience to set in. This undermines the ranges impact and strongly diminishes its potential.

Systems must be in place that allow users to register for access on the Cyber-Range and also gives access to a self-service portal that gives feedback and updates relevant to the individuals users own performance and encourages further progress deeper into the Cyber-Range.

Content of challenges within the Cyber-Range must reflect the technologies and their exploits that are implemented in the real world. These are modern digital problems that need a cyber-centric solutions. To

guarantee availability, the back end framework used to govern the Cyber-Range should be beyond the reach of an end user.

Concerns around readily available automated penetration testing tools (Kali Linux, 2014) capabilities should be addressed. Rules defining the boundaries between framework and Cyber-Range targets must be clearly defined to the end user before the exercise commences. These rules need to be monitorable and enforceable.

5. Address challenges of current cyber-range implementations

Current Cyber-Range implementations struggle to achieve horizontal scale. While extra capacity can be added with the addition of back end servers, increasing the exploit potential of the real world components is often not easily realisable. Often the physical nature of a the Cyber-Range itself becomes a barrier to scale and opportunity. It cannot be broken down, and is too delicate and cumbersome to re-locate.

The National Irish Cyber-Range (CyberRange.ie, 2014) is built to a different standard. The Irish Cyber-Range environment named 'Partholón' was built from the ground up in a modular fashion *Partholón* is designed specifically to address challenges which affect a typical Cyber-Range. It is a low cost implementation, highly scalable in all directions, reproducible remotely through the use of downloadable 3d printing templates and electronic blueprints, easy maintenance and levers open-source technologies. Due to it's modular component build it can be de-constructed and reconstructed for transportation and relocation. Back-end services and apis are scalable in all directions. New and existing cyber challenges are hot swappable into and out of the Cyber-Range environment.

6. What is a modular cyber-range?

A modular Cyber-Range allows for the integration of new technologies, maintenance of the existing components, expansion of the physical platform and interconnectivity management with ease.

A modular Cyber-Range reduces development overhead, construction times, software test and new component integration times. It is a responsive platform able to adapt and flex to address new trends in the marketplace. Implementing a modular design requires the specification of a basic design standard that all elements must adhere to for their communication within the Cyber-Range. It also requires a physical connectivity specification to be defined that outlines essential wiring, power and control systems.

The benefits of this approach can be seen in a *Partholón* simplified example that is constructed from a series of Range-Panels. A Range-Panel is a platform that can operate as an independent Cyber-Range or be daisy-chained to create a larger operational environment or scope. Each Range-Panel can be made up of multiple Range-Tiles. A Range-Tile is a technical node that gives input to or responds to output from a back-end server or other Range-Tile. Range-Tiles can be added or removed as budget allows, with technical complexity and physical operation depending on the exploit.

7. Implementing a modular cyber-range

A modular Cyber-Range is just like a network of districts that make a city, it needs good planning with a well considered layout and overarching long term plan. A modular Cyber-Range can be built from a collection of remotely located Range-Panels. This is a very valuable benefit of Modular Cyber-Range design and is why good planning and adherence to standards are required from the beginning.

Unlike a Capture-The-Flag environment a Cyber-Range bridges the gap between the virtual and real world. It is a construct of software and responsive hardware. The software is responsible for handling the inputs from Range-Tiles and passes responses to hardware on the same or an alternate Range-Tile as output. This relationship determines how all interactions across a Cyber-Range are carried out. Understanding how to alter this relationship either through manipulation of input or output is the ultimate goal of a cyber-defender.

To enable the interchangeable elements that make up a modular Cyber-Range a Range-Panel design must be firstly defined. This standard with determine power, connectors, wiring and api constraints. Below is a simple base design to follow for a modular Cyber-Range:

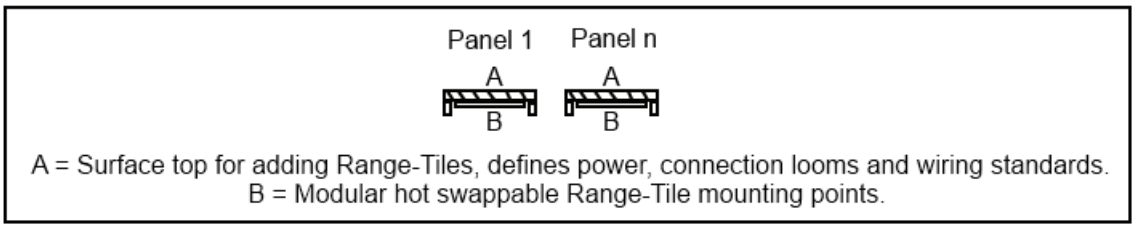

Figure 1: Range-Panel design

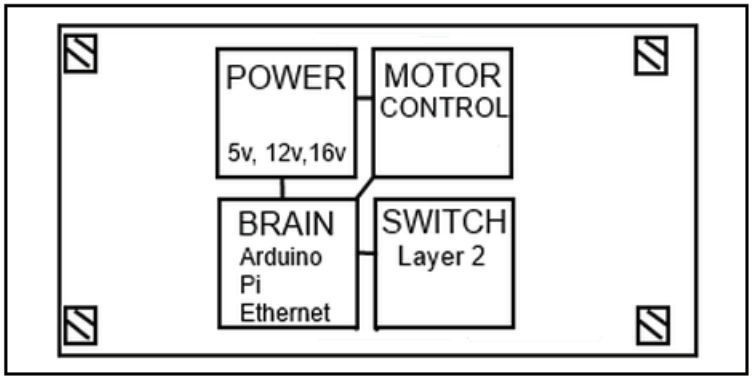

Figure 2: Modular Range-Panel and connecting range tile

In the below example we outline the basics of modular design through a simplified rail service. It demonstrates how Range-Tiles can be linked to perform functions and respond to alternate Range-Tile activities automatically or be overwritten by API input from a Range-Controller.

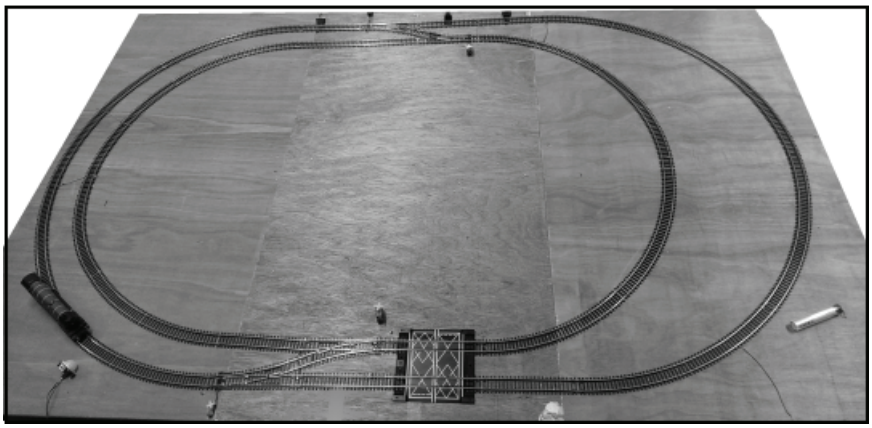

Figure 3: Partholón simplified example

The Range-Tiles built in the above *Partholón* simplified example monitor the trains movements, route decisions and travel data. The train is free to decide the route it wants to take based on safety concerns and other trains on the track. The Range-Controller can alter the trains path by remotely accessing a point system in the track and change the train's path. Manually overriding the trains decisions in this example removes the automated safety measures in place on the Range-Tiles that protect the train and its virtual passengers. The human interaction can orchestrate a systematic failure that will cause the train to ultimately derail and crash.

This simplified *Partholón* example as seen in Figure 4 uses motion sensors, arduino, backend services and servos to influence the course of the train. New Range-Tiles can be added to expand on this functionality.

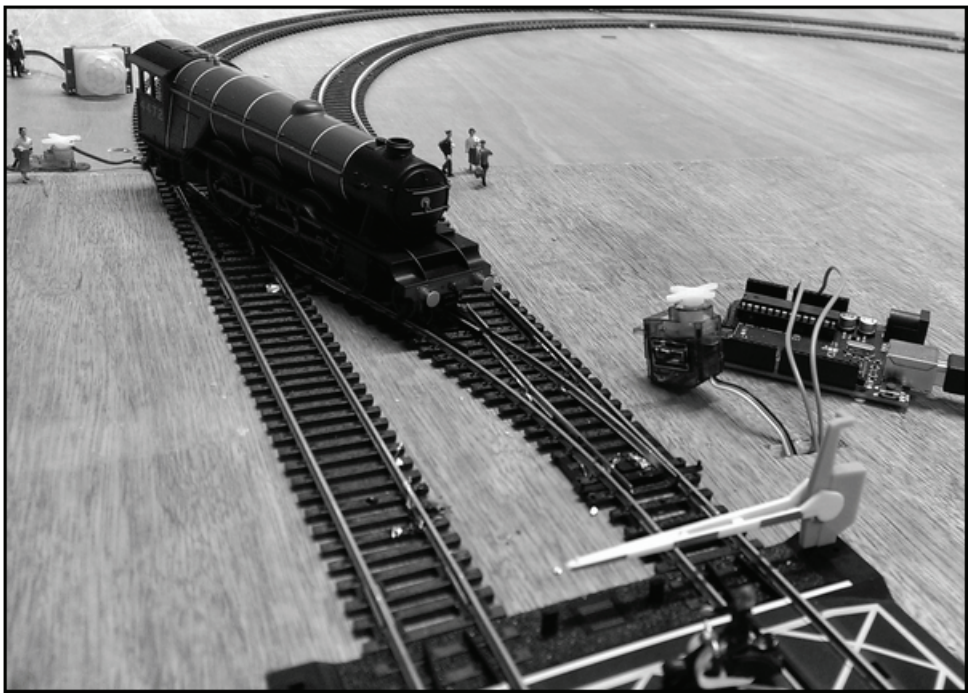

Figure 4: Cyber-Range control devices

8. Management of a modular cyber-range utilization

To create a secure where it must be secure but vulnerable where it can be vulnerable fully scalable environment, three separate networks must be implemented.

- The physical hardware network needs to be able to communicate with the applications network devices Range-Tiles and Range-Panels.
- The users of the Cyber-Range need to be able to access the application network in order to use the Cyber-Range and control the Range-Tiles. They should not be able to influence control on a Range-Panel or gain direct access to a Range-Tile.
- The reporting, maintenance and management services required by a Cyber-Range administrator must be able to contact the Cyber-Panels, Cyber-Tiles, Cyber-Range users and be safe from attack from the users themselves.

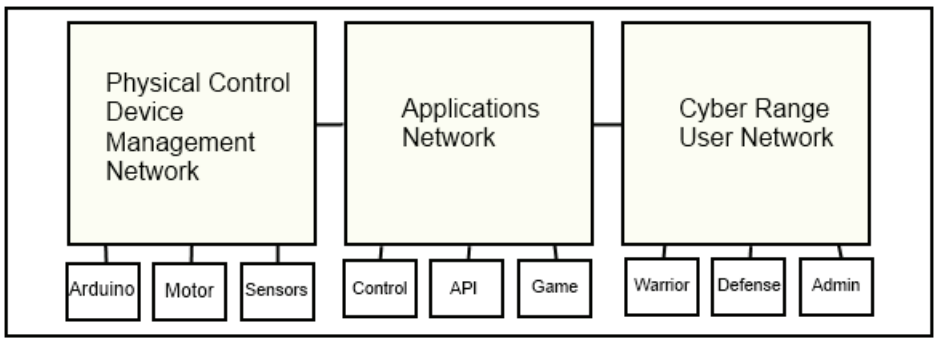

Figure 5: Framework of modular cyber-range

A modular Range-Panel possesses a main control unit, this main unit offers out status information through the API that is consumable by the Cyber-Range admin exclusively.

9. Challenges associated with a modular cyber-range implementation

Adhering to a common structure during construction of Range-Tiles and Range-Panels in a modular Cyber-Range enables the seamless addition of new technologies in an already established environment. It does

however require a willingness to adapt any new technology to an approach that suits integration into the defined modular format chosen by the Cyber-Range owners.

There is a necessity to understand basic electronic circuitry, programming and design fundamentals. Simulating a real world requires each Range-Tile to be able to make it's own decisions but these decisions also need to be reported out to the Cyber-Admin in a secure way.

10. Conclusions

A modular Cyber-Range approach has clear benefits over a standard approach. During the development of 'cyberrange.ie' we found huge synergies that can be made if a standard modular implementation and api existed. It is hoped that *Partholón* will help to define this open standard. The work to date is bringing us in very interesting and new directions. Determining the motivation and cadence behind the attacker and the cyber defender requirement to up skill in the perpetual cat and mouse game is an interesting puzzle. The system we are building will help to answer questions around skill set, protection methodologies, risk management, attack profiling and cyber education. We also believe as the work continues it will prompt us to look for answers to currently unknown questions.

Acknowledgements

We would like to thank the volunteer members of IRISS, OWASP, Engineers Ireland, The Honeynet Project and the staff of The Institute of Technology Blanchardstown for their continued assistance in making the Cyber-Range possible and giving us an audience to test our systems against.

References

Arthur, C. (2014), "Cyberwar heats up with Pentagon's virtual firing range", [online], theguardian.com, http://www.theguardian.com/technology/2011/jun/17/pentagon-virtual-firing-range.

Janssen, C. (2014) "What is a Cyber Range? - Definition from Techopedia", [online], Techopedia, http://www.techopedia.com/definition/28613/cyber-range.

DARPA. "The National Cyber Range: A National testbed for Critical security research",[online], DARPA - Strategic Technology Office, http://www.whitehouse.gov/files/documents/cyber/DARPA%20-%20NationalCyberRange_FactSheet.pdf.

De Montfort University. (2014), "Services",[online], De Montfort University Leicester, http://www.dmu.ac.uk/research/research-faculties-and-institutes/technology/cyber-security-centre/services.aspx.

Department of Defense. (2011) "Department of Defense Strategy for Operating in Cyberspace", [online], Department of Defense - United States of America, http://www.defense.gov/news/d20110714cyber.pdf.

McDuffie, E.L. et all. (2012) "National Initiative for Cyber Security Education Strategic Plan", [online], NICE, http://csrc.nist.gov/nice/documents/nicestratplan/nice-strategic-plan_sep2012.pdf.

FTSI. (2012) "Helping our Wounded Warriors and our Nation by Building Technical Cyber Capabilities", [online], Federal IT Security Institute, http://www.w2cca.org/images/W2CCA-Final.pdf.

Winter, H. (2012) "System Security Assessment Using A Cyber Range", [online], Northrop Grumman Information Systems Europe, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6458960.

Kali Linux. (2014), "Kali Linux - Penetration Testing Redefined", [online], http://www.kali.org/about-us/.

Pridmore, L., Lardieri, P. and Hollister R. (2010) "National Cyber Range (NCR) Automated Test Tools: Implications and Application to Network-Centric Support Tools", [online], Lockheed Martin Global Training and Logistics, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5613581.

Rosenstein, M. and Corvese, F. (2012) "A Secure Architecture for the Range-Level Command and Control System of a National Cyber Range Testbed", [online], Applied Visions Inc., Secure Decisions Division, https://www.usenix.org/system/files/conference/cset12/cset12-final1.pdf.

Microsoft. (2014), "What is a botnet?", [online], Microsoft, http://www.microsoft.com/en-gb/security/resources/botnet-whatis.aspx.

Fujii, N. (2014), "New cyber range promotes cyber security" ,[online], Ka Leo O Hawaii: News, http://www.kaleo.org/news/new-cyber-range-promotes-cyber-security/article_fea371f2-fd7b-11e2-8312-0019bb30f31a.html.

openSCADA. (2014) "We are the good guys", [online], openSCADA, http://openscada.org/.

Andel, T.R., Stewart, K.E. and Humphries, J.W (2014) "Using Virtualization for Cyber Security Education and Experimentation", Air Force Institute of Technology, Proceedings of the 14th Colloquium for Information Systems Security Education, 7-9 June.

CyberRange.ie. (2014) "The Irish National Cyber Range", [online], cyberrange.ie http://cyberrange.ie

A Cyber Attack Evaluation Methodology

Kosmas Pipyros[1], Lilian Mitrou[1, 2], Dimitris Gritzalis[1] and Theodore Apostolopoulos[1]
[1]Athens University of Economics and Business, Athens, Greece
[2]University of the Aegean, Samos, Greece
pipyrosk@aueb.gr
l.mitrou@aueb.gr
dgrit@aueb.gr
tca@aueb.gr

Abstract: Following the identification on an international basis of cyberspace as a new 'domain of warfare', it has become widely (though not fully) accepted that the traditional rules of International Humanitarian Law are also applicable to Computer Network Attacks (CNAs). Despite the fact that there has been considerable progress at the European and International level towards the development of National Cyber Security Strategies and the adoption of an effective comprehensive legal framework of prevention measures against cyber attacks, there is confusion regarding the application of these rules. More specifically, it has not been clarified: a) in which cases do cyber attacks constitute a 'threat or use of force' so that the prohibition of article 2(4) of the UN Charter can apply, b) in which cases do cyber attacks constitute a 'threat to the peace, breach of the peace, or act of aggression' so that the Security Council may decide upon measures to restore international peace and security under Article 42 of the UN Charter, and c) in which cases cyber attacks can be treated as an 'armed attack', making it possible for a UN Member State to respond by exercising its legitimate right of self-defense under Article 51 of the UN Charter. The difficulty in applying the traditional rules of International Humanitarian Law to categorize cyber attacks stems from a number of factors. The most important of them is the failure to estimate properly the impact of a cyber attack in the host country and in the international environment. Additionally, the inability to positively identify the key actor of an attack makes it often quite hard to handle the issue of 'attribution'. The aim of this paper is to propose a model for detecting the effects of cyber attacks and for enabling their categorization on the basis of their type and intensity. The above method requires the identification of the Critical Information and Communication Infrastructures of each State and their ranking in terms of their intensity and seriousness.

Keywords: cyber warfare, computer network attack (CNA), Information and communication systems and technologies (ICTs), critical infrastructures, international humanitarian law (IHL)

1. Introduction

The rapid development of Information and Communication Technologies (ICTs) over the last decades has contributed a lot to the progress of society. The presence of new technologies in every aspect of human life has extended to such a degree that major public sector industries, such as National Security, Education, Government, Health, Public Safety, as well as sectors such as Nutrition, Energy, Economics and Transportation & Communication, are closely related to, if not dependent on new ICTs.

As information and communication systems and technologies are connecting through cyberspace, in order to provide a society's proper functioning and the well-being of its citizens, it's more than clear on the one side that there is an interdependency between cyberspace and new ICTs and on the other side that cyberspace plays an important role acting as the connecting link between them. But what we really mean when we refer to cyberspace?

Although there is no universal definition of cyberspace one could adopt the definition proposed by the US Department of Defense Strategy for operating in cyberspace. This definition which focuses mainly on cyber security issues, states that cyberspace is defined as 'an interdependent and interrelated infrastructural IT network, including the internet, telecommunication networks, computer systems and the systems managing production processes and control in strategic sectors connected to national security' (Nowak, 2013: 7). However, taking into consideration the widespread and growing use of social media one cannot overlook the fact that cyberspace is defined more by the social interactions involved rather than its technical implementations. It is a domain that is becoming more and more a communication channel of information exchange between people functioning in accordance with formal rules, legal regulations in use in the territories of particular countries and operating thanks to the connection of technical resources located on the territory of every single country (Morningstar and Farmer, 2003: 664-667; Nowak, 2013: 9).

The inability to deal effectively with cyber attacks, although they represent a direct threat to a state's ICTs, stems from a number of factors. The most important of them is the failure to estimate properly the impact of a cyber attack in the host country and in the international environment. For that reason, in this paper we propose a cyber attack evaluation methodology, under the specific legal branch of International Humanitarian Law (IHL), for detecting the effects of cyber attacks and for enabling their categorization on the basis of their type and intensity.

2. Real life cyber incidents from the perspective of international law

The advances in ICTs go hand in hand with the first cyber-attack incidents that become more and more sophisticated with the passing of time. The first cyber incidents to be regarded of a military nature were those that emerged during the Kosovo era involving conflicts conducted by non state actors i.e. by the so-called 'patriotic hackers', who seemed however to act under the umbrella of the respective national governments. These types of conflict were characterized '…as the first war on the Internet, in recognition of not only the cyber-attacks but also the broader role played by the Internet, especially in the dissemination of information about the conflict' (Berson and Denning, 2011: 13).

In terms of wide range attacks, the leading one took place in April 2007 in Estonia. That cyber attack was directed against Estonia's critical ICTs leading to the destabilization of the country's financial system and threatening its national security, that is, the ability of the national authorities to ensure the functioning of the basic state fields and to protect the quality of life of its citizens by reducing risks and preventing all kinds of threats to its interests (Tikk, Kaska and Vihul, 2010: 22-23). The Estonia attack was followed by a number of large-scale cyber incidents such as the 'hit' against Georgia, following the increase in intensity of the political conflict between Georgia and Russia. That assault was based mainly on the launching of Distributed Denial of Service (DDoS) attacks against the country's information infrastructure and led to the defacement of the country's public websites (Bumgarner and Borg, 2009: 5-6). The aforementioned aggressions as well as the persistent attacks on U.S ['Operation Aurora' (Zetter, 14.01.2010), 'Ghostnet' (Kassner, 13.04.2009) and DDoS attacks against the New York Stock Exchange (Roberts, 27.09.2012)], Iran [the recent sabotage against Iran's nuclear program with the 'Stuxnet' computer worm (Farwell and Rohozinski, 2011: 23-40; Virvilis and Gritzalis, 2013: 249-250)] and South Korea [aggressions that took place in 2013 and paralyzed three TV stations and part of the country's banking system (Sang-Hun, 20.03.2013)] clearly demonstrate the fact that cyber warfare is a phenomenon that is increasingly relevant (Virvilis, Gritzalis and Apostolopoulos, 2013: 396-403). At the same time, the growing number of cyber events reported on a regular basis has transformed 'Cyberspace' into a battlefield, bringing to light 'Cyber warfare' as 'the fifth domain or warfare' after land, sea, air and space (Lynn, 2010). In parallel, all these incidents brought about a series of discussions over the issue of Computer Network Attacks (CNAs) and their eventual political, economical and social impact on the host State of a Cyber attack but also the international impact regarding this new kind of warfare and its consequences in/for the global strategic environment.

Following that, the critical question that has arisen is whether cyber attack incidents should be met by employing the traditional international law rules in force, or whether they should be considered as something completely different, asking for the introduction of new legislation – new agreements on an international/multinational level. For example, on the one hand Russia, China and other countries favor an international treaty, similar to those agreed on chemical weapons, and have pushed for such an approach to regulating cyberspace. On the other hand the U.S and the EU have repeatedly resisted proposals for an international treaty (O'Connell, 2012: 206). As a matter of fact, despite the opposing viewpoints on the subject according to which 'cyber space is a new military domain and must be understood in its own terms' (Libicki, 2009: 8), it has become widely, though not fully, accepted that the traditional rules of International Law apply also to Computer Network Attacks (CNAs). Besides, all recent institutional documents at European and International level share the same view.

More specifically, the prevailing EU and NATO members' view is that international law suffices to handle issues relating to cyberspace operations. In fact a number of official (customary or treaty based) documents confirm this view. For example, the International Group of Experts (a group of distinguished international law practitioners and scholars) involved in the production of 'the Manual of the International Law applicable to Cyber Warfare' (from now on Tallinn Manual), which was a project launched in the hope of bringing some degree of clarity to the legal issues surrounding cyber operations and which unanimously adopted 'Rules'

meant to reflect customary international law, rejects any characterization of cyberspace as a separate domain calling for its handling by a distinct body of law. On the contrary, the International Group of Experts unanimously has come to the conclusion that the general principles of international law should apply also to cyberspace (Tallinn Manual, 2013: 19). Similarly, at the European level the European Commission, together with the High Representative of the Union for Foreign Affairs and Security Policy, published, on February 2013, a proposal for a cyber security strategy, followed by a draft directive, which aimed to address the issue of Network and Information Security (NIS) and which highlighted that 'the EU does not call for the creation of new international legal instruments for cyber issues' and that 'the legal obligations enshrined in the International Covenant on Civil and Political Rights, the European Convention on Human Rights and the EU Charter of Fundamental Rights should be also respected online' (JOIN, 07.02.2013: 15). The same text, in another point, resumes that 'if armed conflicts extend to cyberspace, International Humanitarian Law and, as appropriate, Human Rights law will apply to the case at hand' (JOIN, 07.02.2013: 16). This same view was reflected as early as 2011, in the U.S International Strategy for Cyberspace where it was clearly stated that 'the development of norms for State conduct in cyberspace does not require a reinvention of customary international law, nor does it render existing international norms obsolete. Long-standing international norms guiding State behavior—in times of peace and conflict—also apply in cyberspace' (The White House, 2011: 9).

Moreover, Rule 10 of the Tallinn Manual, based on article 2(4) of the United Nations Charter, entitled 'Prohibition of the use of force' notes that 'a cyber operation that constitutes a threat or use of force against the territorial integrity or political independence of any State, or that is in any other manner inconsistent with the purposes of the United Nations, is unlawful' (Tallinn Manual, 2013: 45). Nevertheless, this rule does not specify in which cases cyber operations can be considered as attacks that rise to the level of a 'use of force' calling thus for the application of the prohibition of article 2(4) of the UN Charter (extended to Rule 10 of the Tallinn Manual). A potential answer to this question could be given by the next Rule of the Tallinn Manual, ie. Rule 11 stating that 'a cyber operation constitutes a use of force when its scale and effects are comparable to non-cyber operations rising to the level of a use of force' (Tallinn Manual, 2013: 47). It is therefore understood that in order for a cyber operation to be characterized as a 'use of force' a parallel result logic is being employed, meaning that an effort is being made to identify cyber operations that are equivalent in terms of their results to other actions, kinetic or not, that would be described, in conventional terms, as 'uses of force'.

Based on the same logic, and following article 51 of the United Nations Charter, Rule 13 of the Tallinn Manual entitled 'Self-Defence against Armed Attacks' states that 'a State that is the target of a cyber operation that rises to the level of an armed attack may exercise its inherent right of self-defence. Whether a cyber operation constitutes an armed attack depends on its scale and effects' (Tallinn Manual, 2013: 53). However, in this case also, it's not clear in which cases cyber attacks meet the scale and effects requirements so that they can be regarded, classified and handled as an 'armed attack', allowing a UN Member State to respond by exercising its legitimate right of self-defense, under article 51 of the UN Charter. So it can be understood that in both Rule 11 and Rule 13 of the Tallinn Manual, the term 'scale and effects' is a shorthand term that refers to those quantitative and qualitative criteria that should be analyzed in order for someone to be able to determine whether a cyber operation qualifies as a 'use of force' or 'an armed attack'.

3. Scale and effects analysis

The 'scale and effects' concept, which was initially introduced in the so-called Nicaragua Judgment of the International Court of Justice (June 27, 1986) in the 'Case concerning military and paramilitary activities in and against Nicaragua', refers to a set of criteria that gather the qualitative and quantitative characteristics for determining whether or not, a hostile act rises to the level of 'use of force' or to the level of 'armed attack'.

In that Nicaragua Judgment, the International Court of Justice identified the 'scale and effects' criteria as those qualitative and quantitative elements that help differentiate an 'armed attack' from 'a mere frontier incident' (Westlaw, 2007: 84). More specifically, the International Court of Justice noted the need to 'distinguish the most grave forms of force (those constituting an armed attack) from other less grave forms', but chose to give no further details on the subject at hand. As a result, the parameters relating to a clear detection of the 'scale and effects' criteria have not been further identified apart from the indication that they need to be grave.

Therefore, the question remains in relation to the specification of the criteria required to identify which cyber attacks qualify as 'use of force' and, by extension, in relation to the handling of those cases that do not meet the necessary criteria to qualify as 'use of force'.

Taking into consideration that the United Nations Charter does not provide any criteria for determining when an act amounts to a 'use of force', the International Group of Experts adopted an interpretation according to which the critical element for identifying an attack as 'use of force' or as 'armed attack' is the breadth of the impact of this attack. More specifically, they concluded that a cyber operation shall amount to a 'use of force' or to an 'armed attack', if its impact is analogous to the one resulting from an action otherwise qualifying as a kinetic armed attack. By this logic, any attack producing similar results to the ones generated by an attack with the use of conventional weapons, resulting thus in death or destruction, shall meet the requirements of the 'scale and effects' criteria.

Although, the International Group of Experts acknowledged the existence of a legal gap in relation to the identification of the exact point at which an event such as death, injury, damage, destruction or suffering caused by a cyber operation, fails to qualify as an 'armed attack', they were assertive as to what does not qualify as an 'armed attack', namely 'acts of cyber intelligence gathering and cyber theft, as well as cyber operations that involve brief or periodic interruption of non-essential cyber services' (Tallinn Manual, 2013: 55).

Taking thus for granted the fact that the law is unclear as to the characterization and evaluation of a number of cyber attacks, especially in the case of 'use of force' whose impact is not immediately visible, and taking into account the total absence of an institutional framework for the evaluation of the 'use of force' and 'armed attack' concepts in cyberspace, the International Group of Experts proceeded to the adoption of an approach [following Schmitt's consequence-based approach (Schmitt, 1999: 17-19)], that aims to identify, in an objective way, the likelihood of classifying a cyber operation as a 'use of force'.

This approach focuses on recognizing the impact of cyber attacks and on equating it to the corresponding impact caused by other actions (non-kinetic or kinetic) that the international community would describe as 'uses of force'. In these cases, the parallelism and the subsequent analogous treatment of conventional operations, that verge on being characterized as 'uses of force', with corresponding cyber operations that meet the 'scale and effects' requirements, will be the outcome of the evaluation of a number of non exclusive criteria (factors) based on a case-by-case assessment. These criteria (factors) are 'severity' (severity of attacks), 'immediacy' (the speed with which consequences manifest themselves), 'directness' (the causal relation between a cyber attack and its consequences), 'invasiveness' (the degree to which a cyber operation interferes with the targeted systems), 'measurability of the effects', 'military character of the cyber operation', 'extent of State involvement' and 'presumptive legality' (acts not expressly prohibited by international law). Nevertheless, it should be kept in mind that, as the International Group of Experts have clearly clarified, these factors cannot be considered as formal legal criteria.

4. A cyber attack evaluation methodology

As it can been understood, the characterization and categorization of cyber attacks depends largely on the size of their consequences. In other words, the categorization of this type of attacks lies heavily on their impact level both in terms of loss of human lives and in terms of destruction of critical infrastructures. So, the degree of the visible as well as the long-term effects of a cyber attack constitute a critical factor for its categorization and the greater the degree of impact of a cyber attack the more the chances to be characterized as a 'use of force', or even worst, as an 'armed attack' when its size is so great as to cause loss of human lives.

So the critical issue here is the method of measurability of the impact of cyber attack. Unfortunately, as it has already become apparent, the relevant criteria proposed by the International Group of Experts have failed to accurately identify the precise extent of impact of a cyber attack, since its effects are often not readily visible on the short hand and the measurability of the effects of a cyber attack is frequently a matter of subjective interpretation. If the impact level of cyber attacks could be determined through the use of qualitative and quantitative criteria, it would be possibly much easier to classify and categorize them based on the principles of International Humanitarian Law.

On the other hand, one can easily notice that the same impact factors proposed by the International Group of Experts for the categorization and characterization of cyber attacks are also employed as criteria in risk criticality analysis methodologies to prioritize assets and infrastructures. For example, at the European level, the Council Directive 2008/114/EC of 8 December 2008 'on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection', following a relevant European Commission Communication (COM 786, 07.06.2007), identified the following criteria as the minimum set of criteria that should be considered by member states when attempting to assess their critical infrastructures: (i) public safety–including issues such as population affected, loss of life, medical illness, serious injury, evacuation, (ii) economic effect – which takes into consideration the GDP effect, the significance of economic loss and/or the degradation of products or services, (iii) environmental effect – i.e. effect on the public and the surrounding environment, (iv) interdependency – which has to do with interdependencies between critical infrastructure elements, (v) political effects – that is, confidence in the government and (vi) psychological effects – i.e. psychological effects on the population. The evaluation of these criteria takes place in terms of their scope (local, regional, national and international) and time (during and after the incident) (Theoharidou, Kotzanikolaou and Gritzalis, 2009: 35-49).

Respectively, at the international level, the U.S. National Infrastructure Protection Plan identifies the following criteria for evaluating consequences: (i) public health and safety – including their effect on human life and physical well-being, (ii) economic – which takes into consideration direct and indirect economic losses (iii) psychological – i.e. their effect on public morale and the degree of confidence of the people in economic and political institutions and (iv) governance/mission – which related to the effect on the ability of the government or industry to maintain order, deliver essential services, ensure public health and safety and carry out national security – related missions (U.S. Department of Homeland Security, 2009).

It becomes clear that the evaluation criteria used for assessing critical ICTs, are focused more on evaluating risks related to external impacts that is, impacts associated with socioeconomic consequences and their effect on citizens, since they are directly linked to the critical infrastructures affected per se and indirectly associated to the implications of the collapse or degradation of these critical ICTs for the well being of the citizens. This approach comes in contrast to the traditional risk analysis methodologies that focus more on the implications of the collapse or deterioration of infrastructure in the respective department or agency that relates to it (internal impacts), rather than on the external impacts of this collapse or deterioration to the citizens. Consequently, there is a strong interdependency between cyber attacks and the corresponding risk criticality analysis methodologies used for assessing critical infrastructure and networks (ICTs) since in both cases are characterized by the same the impact factors during their evaluation process.

Based on everything mentioned above, one could proceed to an assessment of cyber attacks by adopting risk-based criticality analysis methodologies. A case in point is the generic risk-based criticality analysis methodology proposed by Theoharidou, Kotzanikolaou and Gritzalis (2009: 40-47) by which, a detailed list of impact criteria is presented for assessing the criticality level of infrastructures. What differentiates this method from traditional risk analysis methodologies is the fact that it assumes the same societal and sector-based impact factors used by the International Group of Experts for characterizing and assessing the intensity of cyber attacks, allowing thus the parallelism and the adoption of the same evaluation criteria for assessing cyber attacks.

This criticality analysis methodology, whose primary role is to be used as a base for assessing risk associated with critical ICTs, can also serve as a scale for measuring the intensity of cyber attacks in order to enable a quantification of the 'scale and effect' criteria, using qualitative and quantitative variables such as the ones recommended by the International Group of Experts in the Tallinn Manual, and possible adopting other criteria, so that it can become easier to identify when such acts verge on the so-called 'use of force' standard, which is used for determining whether or not a State has violated Article 2(4) of the United Nations Charter and the related customary international law prohibition.

Furthermore, the same methodology could be used to indicate whether a cyber operation comes to the level of being characterized as a 'use of force' or as an 'armed attack' allowing thus a UN Member State to respond by exercising its legitimate right of self-defense according to Article 51 of the UN Charter. Similarly, the above method could serve as a scale for the Security Council to decide when a cyber attack constitutes 'threat to the peace, breach of the peace or act of aggression', so that the required measures to restore international peace

and security under Article 42 of the UN Charter can be adopted. In other words, the adoption of the criticality risk analysis methodology can serve as a means for estimating the impact of a cyber attack in the host country and in the international environment.

Conclusively, the discussed evaluation methodology, using as a reference point the above mentioned criteria, could be used as a method for stressing areas where there is uncertainty or disagreement in an number of legal analyses, and for making available a means for addressing all issues having to do with 'use of force'. In addition, this methodology can act as a basis for the assessment and classification of cyber attacks that are intended towards software systems that may constitute a component of a critical infrastructure.

Our future work will be focused on a more accurate approach of the above mentioned cyber attack methodology in order to determine and evaluate the impact factors of a cyber attack on the basis of their type and intensity, for enabling their categorization under the principles of International Humanitarian Law.

References

Barmpatsalou, K., Damopoulos, D., Kambourakis, G., Katos, V. (2013) 'A critical review of 7 years of Mobile Device Forensics', *Digital Investigation*, vol.10, no. 4, December, pp. 323-349.

Berson, T. A. and Denning, D. E. (2011) 'Cyberwarfare' , *IEEE Security & Privacy*, vol. 9, no. 5, September/October, pp. 13-15.

Council Directive (EC) 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection, OJ L 345/75.

Department of Homeland Security, (2009) *National Infrastructure Protection Plan*, Washington: US Department of Homeland Security.

European Commission, Communication from the Commission of 12 December 2006 on a European Programme for Critical Infrastructure Protection, COM (2006) 786 Final, Brussels, Belgium, 12.12.2006.

European Commission, Joint Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Cyber security Strategy of the European Union: An Open Safe and Secure Cyberspace', JOIN(2013) 1 Final, Brussels, Belgium,7.2.2013.

Farwell, J. and Rohozinski, R. (2011) 'Stuxnet and the Future of Cyber War', *IISS, Survival: Global Politics and Strategy*, vol. 53, no. 1, pp. 23-40.

Kassner, M. (2009) *Ghostnet: Why it's a big deal*, [Online], Available: http://www.techrepublic.com/blog/it-security/ghostnet-why-its-a-big-deal/1339/ [2 Oct 2013].

Libicki, M. (2009) *Cyber deterrence and cyber war*, Santa Monica: The Rand Corporation.

Lynn, W. (2010) 'Defending a New Domain: The Pentagon's Cyberstrategy', *Foreign Affairs,* vol. 89, no. 5, Available: www.foreignaffairs.com/articles/66552/william-j-lynn-iii/defending-a-new-domain [2 Oct 2013].

Morningstar, C. and Farmer, F. R. (2003) 'The Lessons of Lucasfilm's Habitat*:* The New Media Reader', Wardrip-Fruin and Nick Montfort (ed.), Cambridge: The MIT Press.

Mylonas, A., Kastania, A., and Gritzalis, D. (2013) 'Delegate the smartphone user? Security awareness in smartphone platforms', *Computers & Security*, vol. 34, pp. 47-66

Mylonas, A., Meletiadis, V., Mitrou, L., and Gritzalis, D. (2013) 'Smartphone sensor data as digital evidence', *Computers & Security (Special Issue: Cybercrime in the Digital Economy)*, vol. 38, pp. 51-75.

O'Connell, M. (2012) 'Cyber security without Cyber War', Journal of Conflict & Security Law, vol. 17, no. 2, pp. 187-209, Available: http://jcsl.oxfordjournals.org/content/17/2/187.full.pdf?keytype=ref%2520&ijkey=T6J6KDRCRcHM4Ao [2 Dec 2013].

Roberts, P. (2012) *Leading US banks targeted in DDoS attacks*, [Online], Available: http://nakedsecurity.sophos.com/2012/09/27/banks-targeted-ddos-attacks/ [2 Oct 2013].

Sang-Hun, Ch. (2013) *Computer Networks in South Korea are paralyzed in Cyber attacks*, [Online], Available: http://www.nytimes.com/2013/03/21/world/asia/south-korea-computer-network-crashes.html?pagewanted=all&_r=0 [2 Oct 2013].

Schmitt, M. (1999) 'Computer Network Attack and the Use of Force in International Law: Thoughts on a normative framework', *Columbia Journal of Transnational Law*, vol. 37, Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1603800.

Schmitt, M. (ed.) (2013) *Tallinn Manual on International Law Applicable to Cyber Warfare*, Cambridge: Cambridge University Press.

The White House, (2011) *International Strategy for Cyberspace: Prosperity, Security and Openness in a* Networked *World*, Washington: The White House Publications.

The United States Cyber Consequences Unit (US-CCU), (2009) *Overview by the US-CCU of the Cyber Campaign against Georgia in August 2008*, US-CCU Special Report.

Theoharidou, M., Kotzanikolaou, P., and Gritzalis, D. (2009) 'Risk-based criticality analysis', *Critical Infrastructure Protection III*, Springer, pp. 35-49.

Tikk, E., Kaska, K. and Vihul, L. (2010) *International Cyber Incidents: Legal Considerations*, Tallinn: CCD COE.

Virvilis, N., and Gritzalis, D. (2013) 'The big four - What we did wrong in advanced Persistent Threat detection?', Proc. of the 8th International Conference on Availability, Reliability and Security, Germany, pp. 248-254.

Virvilis, N., Gritzalis, D., and Apostolopoulos, T. (2013) 'Trusted Computing vs. Advanced Persistent Threats: Can a defender win this game?', Proc. of the10th IEEE International Conference on Autonomic and Trusted Computing, Italy, pp. 396-403.

Westlaw (2007), Case Concerning Military and Paramilitary Activities in and against Nicaragua, [Online], Available: http://www.ilsa.org/jessup/jessup08/basicmats/icjnicaragua.pdf [2 Oct 2013].

Zetter, K. (2010) *Google Hack Attack was Ultra Sophisticated, New Details Show*, [Online],
Available: http://www.wired.com/threatlevel/2010/01/operation-aurora/#ixzz0deHCunGn [2 Oct 2013].

The Changing Character of war in the Global Information Age

Anthimos Alexandros Tsirigotis
The University of Reading, Reading, UK
a.tsirigotis@pgr.reading.ac.uk

Abstract: This paper presents the methodological approach of an ongoing research which has been inspired by the big question of whether or not the cyber weapons change the character of war. The literature tackles this issue mainly by trying to give 'yes' or 'no' answers to the frenzy of questions about the probability of a cyberwarfare scenario and whether or not the traditional weapons are becoming obsolete. The mainstream of academia uses empirical data for arguing that the reasons cyber weapons have attracted the public interest are more due to their newness than by virtue of any value they have in military affairs. 'War as always' succinctly summarises their arguments. Nevertheless, this research work argues that it is not in the empirical where the changes should be traced. It is, instead, in the discourse on war that the cyberspace brings about changes. In line with critical war studies, it is founded on the principle that the ontology of war is not reducible to the fighting power manifested in battlefields and that it has also a sociopolitical dimension. Critical realism offers the theorisation of how discourse creates the social reality within which war is taking place. In each era, social agents work within a network of social practices for giving meaning to the social world by means of *physical elements, sociological elements, cultural/psychological elements, and discourse.* In this work, the object of research is constructed within *discourse* which is developed in the social practice that concerns the revision of the British national defence and security policy. The core research question to probe is how language patterns and meanings have changed from 1998 to 2013 so as for the cyberwarfare discourse to emerge in this British policy making context. Accordingly, it uses a corpus of policy documents processed according to the Modern Diachronic Corpus-Assisted Discourse Studies. The process starts off from meticulously examining the use of specific lexical items throughout this period of time and then it focuses on how they have been used within the textual context. By doing so, the purpose is to unearth how new lexical patterns which bear new meanings have been emerged. Accordingly, it is offered the genealogy of the cyberwarfare discourse which, however, needs further research in order to prove how it finds hegemonic place within social reality, how it is recontextualised in other social practices and, finally, how it is operationalised.

Keywords: cyberspace, cyberwarfare, critical war studies, corpus-assisted discourse analysis

1. Introduction

The swift and abrupt introduction of the cyber capabilities into military affairs has created conflicting arguments concerning their potential for changing the character of war. The stabbing question of whether or not cyber war is a probable scenario has become a tug-of-war between those who consider that it 'will take place' and those who reject it. 'Cyber' has become the buzz word of at least the last decade being added as a prefix to practically every human activity. Upon the explosion of the Internet, fraud was instantly converted to 'cyber fraud', crime to 'cyber-crime', weapons to 'cyber-weapons' and war to 'cyber-war'. The dynamics of the new medium was as influential as it self-imposed its lexicon on a wide series of activities. This practice was also adopted by social scientists who argued that the world is heading to a new era during which the Information Communication Technologies have a catalytic role to play in 'sanitising' societies.

In war studies, Roennfeldt contends that the despotic influence of Clausewitz does not leave enough space for a new interpretation of the phenomenon of war (Roennfeldt 2011). Meanwhile, some flamboyant titles such as 'Preparing for a digital Pearl Harbour' (Schwatz 2007) and 'Preparing for a Digital 9/11' (Greenberg 2012) have created a war atmosphere in the cyber domain lending credence to the term 'cyber war'. The often cited cases of Estonia (2007), Georgia (2008) and the Stuxnet (2010) came to further raise these aspirations. The arguments found in the relevant literature are quite controversial and a succinct overview of them is offered in a recent issue of the *Journal of Strategic Studies*. There, on the one hand, academics with an interest in war and in strategy raised their voices against the abuse and misuse of the term 'war' for describing a series of malicious activities through cyberspace and they explain 'Why the Sky Will Not Fall' (Gray 2013). On the other end of the spectrum, an emerging school of thought argues that the cyberspace and the way that it changes the character of war deserve more research. For them, the new political context raises the need for a more in-depth discussion concerning the way in which war is conceptualised or how power is now defined.

This paper presents a research design that is informed by both arguments, but which chooses to follow a different way as far as its philosophical stance and methodology are concerned. Its objective is to examine the changes in the perception of the British military and security elites of the character of war in the Global Information Age. Accordingly, the following sections firstly present the two conflicting arguments found in the

literature which are summarised here under the titles of the 'Thesis' and 'Antithesis'. Secondly, this paper discusses the *cultural turn* in war studies which puts emphasis more on the discourse on war and it explains how the latter defines the reality of war. The third part offers the philosophical stance of the whole research design, namely it describes how critical realism conceives social life. Lastly, it describes how the methodology of Critical Discourse Analysis (CDA) is applied in this research in an effort to open-up the discussion concerning new research approaches in war studies.

2. The Thesis: 'war' freed of prefixes

In 2011, both Betz and Rid tried to bring the whole debate about the so-called cyber war back on the right tracks by emphatically arguing that 'Cyberwar is not coming' (Betz and Stevens 2011, p.88) and that 'Cyber War Will Not Take Place' (Rid 2011) respectively. The former accentuated the marginal, if any at all, strategic potentials of the cyber war as a political means for the states to impose their will on their adversaries mainly due to the attribution problem. An orphan attack, namely an attack stemming not from an internationally identifiable political entity, is bereft of any power to bring political results to the advantage of its perpetrator. Betz acknowledges, however, the contribution of the cyber means to serve as an accomplishment to the hard power of the kinetic weapons. In essence, this point of view supports that the new capabilities offered in cyberspace will be adjusted to the existing war paradigm and that, as much as revolutionary or spectacular they may seem, ultimately they will be assimilated by the geography of war. This school of thought revokes the literature concerning the sea changes that the air power was supposed to bring for quick and decisive military victories (Rattray 2001) and which, in the end, proved to be wishful thinking. Similarly, Betz and Stevens contend that cyber power does not bring any rupture with the past in the strategic thinking, for 'Technology cannot make up for all the weaknesses of strategy; often what it gives with one hand it takes away with the other' (2011, p.89).

Moreover, Rid rejects cyber war as the next war paradigm because the force that cyber-attacks claim to have is neither violent nor instrumental and political in nature (2011). He rejects the utility of the cyber-attacks because they exert force only through a 'mediated sequence of causes and consequences' (ibid. p.9) and not straightforwardly, let alone that, by now, there have not been any considerable life losses due to any cyber-attack. In 2013, he further accentuated the indirect and unqualified nature of the cyber-attacks by explaining that they lack the political authority needed in order to be converted from sheer disturbance to useful political means of coercion (2013). Thus, they are bound to remain unqualified, or as Liff dubs 'as a destabilizing weapon, although warranted, may be exaggerated' (2012, p.425). In that way, he defends the 'enduring nature of war' (Gray 2010) in the Global Information Age which still remains a mixture of passion, chance and rationality.

In sum, the Thesis defended by strategists and war scholars reiterates that war always remains a political means that involves physical violence for imposing a state's will on its opponents. The practice of adding prefixes before 'war' for indicating the newness in its conduct does not mean that the new type of war brings changes to the nature of war. Accordingly, cyberspace should not be considered ex facie to have war value and to function simply because the prefix 'cyber' happens to be placed before the term 'war'. In contrast, there are insurmountable constraints for the cyber-attacks to have political and strategic value.

3. The Antithesis: Cyber war merits more research and less hype

While the Thesis defends the enduring nature of war, an emerging school of thought proposes that the new reality changes the way that war and military power should be conceptualised. Stone proposes in what sense the attacks through the new medium may serve the war objectives while he acknowledges that many cyber incidents are indiscriminately given the title of 'cyber war'. He rejects Rid's argument that the violence imposed on the targets by means of weapons has to be lethal if it is to have any war value (Stone 2013). Stone argues that the military value and impact of the attacks on the outcome of war, whether cyber or not, should not be evaluated on the basis of how much lethal they are. Instead, it is the violence produced by an attack that puts pressure on the target and which ultimately forces him to comply. He underlines that the Western liberal way of war rests on exactly this idea, namely on the disassociation of force, violence and lethality. In this triangular scheme, Stone believes that cyber-attacks have potentials in the axis of force – violence by playing the role of 'violence multiplier' (2013, p.106) even without exerting any physical force. Within this framework he concludes that cyber-attacks 'could constitute acts of war' (ibid., p.107).

The contribution of the Antithesis to the cyberwarfare debate is, in essence, a new way to speak about power in the Global Information Age, but also about the place that war has in the international system. To start with the first one, there is a budding literature arguing that within cyberspace power should be conceived not in the same way as in the physical world. The actors in the cyber domain become powerful through an immaterial process which relies more on their ability to freely promulgate their own images and their own narratives and less on what assets they possess in the physical world (Kim et al. 2012). Betz and Stevens dub this new demonstration of power as the *Productive cyber power* and they define it as 'The constitution of social subjects through discourse mediated by and enacted in cyberspace, which therefore defines the 'fields of possibility' that constrain and facilitate social action' (2011, p.50).

Roennfeldt, in line with Betz and Stevens, argues that there is a need to reconceptualise war in what he dubs the *Productive war.* The latter is in stark contrast to the compulsory power of the kinetic war and, accordingly, modern political entities become powerful in the 'discursive battlefields' and not in the war battlefields. Roennfeldt proposes the disassociation of power with violence and he argues that 'power now rests on legitimacy, [and therefore] communication has replaced force as the decisive means in power politics' (2011, p.46). These non-territorially defined battlefields are created in what Roennfeldt dubs the 'local centres of power' in which anyone is powerful by '[…] producing the discursive effects that reproduce and modify discursive patterns of domination' (ibid., p.56)[1].

The contribution of Singh further explains the theoretical foundations of cyber power as a distinct kind of power. He argues that the power of ICT should be sought in their ability to frame new identities and to impose new meanings in world politics. This is how he defines the *'meta-power'* which is not traced on the present time but, instead, on precedent time frames. People interact in actual time through cyberspace and, in essence, what they do is to work collectively for defining the conceptual social frameworks of the future. This is why the '*meta power'* precedes the *'instrumental power'*. He explains that 'When nation – states fight territorial wars, similarly the meaning of a nation – state or security understood in territorial terms has been *imagined* through prior interactions' (Singh 2013, p.7).

To continue with how war as a political means is now defined within the international system, the recent work edited by O'Conell offers useful insight (2012). O'Connel chaired the Committee on the Use of Force commissioned by the International Law Association (2005-2010) to report on the definition of war in international law. The main finding of the Committee is that 'war' is an ill-defined term in international law and that it is now replaced by the term 'armed conflict'. It is worth mentioning here the tendency in international law to define more restrictively the exertion of violence in world politics. The term 'war' is now considered to be a rather vague one and, instead, there is a tendency to elaborate a stricter law framework which meticulously describes the exertion of physical violence in international politics.

To conclude, the Antithesis responds to the objections of the cyber sceptics by underscoring that cyberspace and, in essence, the new Global Information Age changes the way people now speak and think about war and military power. Changes, therefore, should not be considered to lay on the empirical manifestations of war but on the discourse on war. The research in hand is about the latter and its objective is to unearth how modern discourse creates new meanings in military affairs. As it is explained in the next paragraphs, the reality of war is constructed by social agents in a different way in every era: its objectives, the identities of the enemies, the threats against the national security but also the technologies used in war operations.

4. The Cultural turn in war studies

In the core of war always lies combat which is inherently violent. Yet its character is reducible neither to its technologies nor to its combats since they are constructed within discourse as well. Technology as a driver for change in the character of war is not considered here to have enough analytical power, for its development does not follow any predefined pattern, but is socially constructed (Jasanoff 2004). The use of the cyber dimension for the purposes of war does not suffice to explain why the character of war is changing. Technology, before taking a specific shape was, in first place, conceived as a meaning in the real world. Change, therefore, does not stem from the technology of weapons, but from the way that societies create the social world and their needs in there.

[1] For further reading see Karatzogianni, A. (2006) *The politics of cyberconflict,* London, Routledge.

This is essentially what lies at the heart of the *cultural turn* and John Lynn explains that 'the essential value of using a cultural approach in military history is precisely in distinguishing the mental from the material' (2003, p.xix). Technological determinism has resulted in reducing war to the fighting power of its weapons, and its epistemology to the analysis of its battles since it is there that the power of weapons is manifested. The discourse of war was considerably neglected until the *cultural turn* was proposed as a new research philosophy in war studies and military history (Barkawi and Brighton 2011, Lynn 2003, Black 2013, Porter 2009). Poststructuralism and critical realism inform the cultural turn in a way that now war studies and military history direct their research efforts more towards the understanding of how discourse creates the reality of war, while acknowledging that war is always about fighting. Barkawi and Brighton underscore this turning point away from the supremacy of battles in war studies and they offer a more thorough view of the ontology of war:

> *It is an ontology that retains the power of war-centered analysis without limiting inquiry to a focus on war fighting. We hold on to the ontological primacy of fighting, but wrest it from the instrumentality its historicity demands.*
>
> *(2011, p.136)*

Now, the changes in the character of war are traced in the realm of ideas, namely in how perceptions and beliefs are constructed and not in the 'capacity for employing force' (Black 2013, p.299). Accordingly, the technology of the weapons, how they deliver their destructive force for the purposes of war, the construction of the identity of enemies, the objectives for deploying armed forces, the ethics of war and even victory and defeat; they are all constructed within discourse.

Cyberspace is, in essence, a technological product which has been developed not in a social vacuum. Instead, its origins lay on the 'imaginative faculties, cultural preferences and economic or political resources' (Jasanoff 2004, p.16) of the present era. Cyberspace was not created in the 90's when the Internet started becoming a mainstream technological product; it was initially created as an idea, as scientists started thinking of intelligible machines that could substitute men. The origin of the idea of cyberspace or, differently, its *myth* is still an embryonic research field. Rid, currently working on this area, proposes that the idea can be traced back to the 40's to the science of cybernetics. What is of interest here is that cyberspace, as any other technological product, 'both embeds and is embedded in social practices, identities, norms, conventions, discourses, instruments and institutions - in short, in all the building blocks of what we term the *social.*' (ibid., p.3). Cyberwarfare, therefore, is not about waging war through a new medium, as was the case when aircraft changed the conduct of war. In this work, cyberwarfare is not reduced to its prefix, namely to 'cyber' or, as the proponents of the *cultural turn* would say, it is not reduced to its fictitious battles which are fought in cyberspace.

The cultural turn has turned the interest towards discourse and to critical realism in order to conceptualise the war phenomenon. The next section describes the philosophical stance on which the discourse on war is founded.

5. Critical realism

Critical realism distinguishes between the natural and the social world. While the former remains a given entity through time, the social world is constructed by people i.e. by social agents. In each era, they work continuously within a network of *social practices* for giving meaning to it under the influence of social structures. Social practices have a mediating function between the abstract social structures and the specific actions-events that take place in social life. They offer a quite stable 'pattern' for the social agents to come together and work for producing these actions-events. These 'habitualised ways, tied to particular times and places, in which people apply resources (material or symbolic) to act together in the world' (Chouliaraki and Fairclough 1999, p.21) are influenced by the social structures which, inasmuch as they are socially constructed, offer only relative stability for a specific period of time and in a specific geographical place.

In this malleable ontology, the network of social practices accomplishes a specific work: it serves as the domain where social praxis is taking place and, at the same time, it changes the world. In specific temporal and spatial dimensions, the changes found in social practices can be traced horizontally, to the rest of the network, and upwards, to social structure. The changes are sought in the three specific functions that social practices have, i.e. in the mode of *production*, *identification* and, finally, *representation* (Chouliaraki and Fairclough

1999, p.21-28). As modes of production, they work for delivering specific products (e.g. educational system, fine arts, national security). By doing so, at the same time they create the social identities of people participating in these practices. The third function is about representation and it can be summarised in that 'people always generate representations of what they do as part of what they do' (ibid., p.26). For instance, military parades offer representations of the work accomplished within the social practice of national defence.

Social practices produce work, give identities and offer representations of what they do by means of specific resources. Fairclough distinguishes among four such categories of resources: *physical elements, sociological elements, cultural/psychological elements, and discourse* (2000, p.168). He calls them 'moments' which are combined all together in order for social practices to perform their functions. Each one of them is for naught unless it is brought together with the others. The objective of social practices is to create synergies or, as Fairclough explains, to create specific 'articulations' among these moments.

In this way, social practices are a sum of moments, each one bearing its discrete semiotic value which, however, remains inoperable and imperceptible out of a specific 'articulation' (combination) of different moments. Discourse, as one of the four moments of social practices, does not refer only to the written language but to:

> *the semiotic elements of social practices. [which] includes language (written and spoken and in combination with other semiotics, for example, with music and singing), nonverbal communication (facial expressions, body movements, gestures, etc.) and visual images (for instance, photographs, film). The concept of discourse can be understood as a particular perspective on these various forms of semiosis - it sees them as moments of social practices in their articulation with other non-discursive moments.*
>
> *(Chouliaraki and Fairclough 1999, p.38)*

Systematic analysis of the written language can, therefore, offer valuable insight in the way that the world is changing. Critical Discourse Analysis (CDA), described in the following section, treats texts not in terms of the language used but, instead, as moments of semiotic value which bear meanings and create social reality as they are produced, disseminated and interpreted within the network of social practices.

6. Critical discourse analysis in war studies

Critical Discourse Analysis was introduced in academia in the early 90's by van Dijk, Fairclough, Kress, van Leeuwen and Wodak as a comprehensive methodology for the systematic examination of discourse; discourse in its full remit, namely as a resource for people to construct the social world. CDA should not be considered as a rigid research method, meaning that it is more about the philosophical stance of the researcher towards social life and change within it, and less about offering a standardised way of data collection and analysis. The objective is always to examine the 'relations between discourse and non-discoursal elements of the social' so as to unearth 'how changes in discourse can cause changes in other elements' (Fairclough 2010, p.357). The way, however, for reaching there is an open-ended one, as explained by Wodak and Meyer (2009, pp.23-33).

This research work aims at examining how cyberwarfare discourse appeared within the social practice that concerns the revision of the British national defence and security policies. As Fairclough (2010, pp.367-371) explains, discourse follows a specific pattern which starts with its emergence within the network of social practices and it ends to its operationalisation in social life (*emergence, hegemony, recontextualisation, operationalization)*. This research is about the genealogy of cyberwarfare discourse, which is created as high rank military officers, politicians, academics, representatives of high influencing think tanks and representatives of the civil community gather together for revising the national defence and security policies of the United Kingdom.

The documents used span from 1998 to 2013, and they were produced by both Houses during the process for the revision of the British Strategic Defence and Security Review (SDSR) and the National Security Strategy (NSS). They describe the action plan of the government for tackling the security threats and for galvanising the national security, each time in reference to the security context. A revised SDSR and NSS signal a significant change in world politics which dictates a new national approach in terms of security. Both texts create the British perception of world politics and of the security context in a specific period of time; they create the identity of the threats; they assign new roles to the armed forces and they create relationships among social

agents each time they found or cease an operational centre. Hence, they are treated here as moments of semiotic value the meanings of which, once they are published, flow among the rest of social practices; they develop synergies with more moments of semiotic value and in the end they change the social structure. In this way, texts are both the trace of a process, but also the cue to a new one, inasmuch as they are constantly subject to different interpretations.

The texts are compiled in a corpus which offers abundant information about how language has been used from 1998, when the lexical item 'cyberspace' firstly appeared in the Security Defence Review, up to 2013. By 2013, two revisions of the so-called Cyber Security Strategy had been published and a lot of written material had been produced (written and oral evidences) during the parliamentary process for the next Defence and Security Review due to be published in 2015. The purpose, accordingly, is to 'track how language patterns and meanings' (Partington et al., p.265) have been stabilised within this corpus in a way that permitted the emergence of the cyberwarfare discourse in British policy making. For doing so, the Modern Diachronic Corpus-Assisted Discourse Analysis (MD-CADS) approach (Partington et al.) is followed. The corpus has been harmonised through annotation and has been converted to XML-valid, TEI conformant corpus, which is considered as the preliminary step of a CADS research design (Cirillo et al. 2007). Word lists, concordance lines and collocates were then extracted for different time periods using the AntConc free software (Anthony 2014). In this way, it was possible to find specific lexical patterns that have been created through time and then to examine how they work in the context of the documents, so as for new meanings to emerge within discourse.

To conclude, contemporary research on how the conduct of war is changing in a world exponentially interconnected has much to offer in so far as it directs its interest towards the discourse on war. Cyberspace brings about changes, for it enables a new discourse to emerge as far as the utility of war, its conduct and the nature of military power are concerned. The ongoing research intends to offer a small particle of knowledge on how this new discourse emerges in the British context. Yet much more research, transdisciplinary in nature, is necessary, so as to have the whole picture of how this discourse is operationalised within social life.

Acknowledgements

I would like to thank my colleagues at the University of Reading for their valuable feedback on my work during the Doctoral Research Seminars organised by the Department of Politics and International Relations. I also need to thank Dr. Andrew Liaropoulos for his guidance during the last three years.

References

Anthony, L. 2014 AntConc (Version 3.4.0) [Computer Software]. Tokyo: Waseda University. Available: http://www.antlab.sci.waseda.ac.jp/.

Barkawi, T. and Brighton, S. 2011 "Powers of War: Fighting, Knowledge, and Critique", *International Political Sociology,* 5**,** 126-143.

Betz, D. and Stevens, T. (2011) *Cyberspace and the state : toward a strategy for cyber-power,* Abingdon, Routledge.

Black, J. (2013) *Introduction to global military history: 1775 to the present day*, Routledge.

Chouliaraki, L. and Fairclough, N. (1999) *Discourse in late modernity : rethinking critical discourse analysis,* Edinburgh, Edinburgh University Press.

Cirillo, L., Marchi, A. and Venuti, M. The CorDis Corpus: Mark-up and Related Issues. Proceedings from the Corpus Linguistics Conference Series, 2007.

Fairclough, N. 2000 "Discourse, social theory, and social research: The discourse of welfare reform", *Journal of Sociolinguistics,* 4**,** 163-195.

Fairclough, N. (2010) *Critical discourse analysis : the critical study of language,* Harlow, Longman.

Gray, C. S. (2010) *The strategy bridge : theory for practice,* Oxford, Oxford University Press.

Gray, C. S. 2013 Making Strategic Sense of Cyber Power: Why the Sky in not Falling. Carlisle, PA: Strategic Studies Institute and U.S. Army War College Press.

Greenberg, K. J. 2012. *Preparing for a Digital 9/11* [Online]. Infowars.com. Available: http://www.infowars.com/preparing-for-a-digital-911/ [Accessed 15 December 2012].

Jasanoff, S. (2004) *States of knowledge : the co-production of science and social order,* London, Routledge.

Karatzogianni, A. (2006) *The politics of cyberconflict,* London, Routledge.

Kim, H. M., Lee, D. and Feiock, C. R. 2012 "Network Power and Militarized Conflicts", *Armed Forces & Society,* 38**,** 291-317.

Liff, A. P. 2012 "Cyberwar: A New 'Absolute Weapon'? The Proliferation of Cyberwarfare Capabilities and Interstate War", *Journal of Strategic Studies,* 35**,** 401-428.

Lynn, J. A. (2003) *Battle : a history of combat and culture,* Boulder, Colo.; Oxford, Westview.

O'connell, M. E. (2012) *What is war? : an investigation in the wake of 9/11,* Leiden, Martinus Nijhoff.

Partington, A., Duguid, A. and Taylor, C. *Patterns and meanings in discourse : theory and practice in corpus-assisted discourse studies (CADS)*.

Porter, P. (2009) *Military orientalism*, Columbia University Press.

Rattray, G. J. (2001) *Strategic warfare in cyberspace,* Cambridge, Mass. ; London, MIT Press.

Rid, T. 2011 "Cyber War Will Not Take Place", *Journal of Strategic Studies,* 35**,** 5-32.

Rid, T. 2013 "More Attacks, Less Violence", *Journal of Strategic Studies,* 36**,** 139-142.

Roennfeldt, C. F. 2011 "Productive War: A Re-Conceptualisation of War", *Journal of Strategic Studies,* 34**,** 39-62.

Schwatz, J. 2007. *Preparing for a digital Pearl Harbor* [Online]. New York. Available: http://www.nytimes.com/2007/06/24/business/worldbusiness/24iht-cyber.1.6299676.html?_r=1& [Accessed 15 October 2012].

Singh, J. P. 2013 "Information Technologies, Meta-power, and Transformations in Global Politics", *International Studies Review,* 15**,** 5-29.

Stone, J. 2013 "Cyber War Will Take Place!", *Journal of Strategic Studies,* 36**,** 101-108.

Wodak, R. and Meyer, M. (2009) *Methods of critical discourse analysis,* London, SAGE.

Masters Research papers

Spymasters Tools: A Comparative Approach to Side Channel Attacks

Tuğçe Kalkavan
Yasar University, Izmir, Turkey
tugce.kalkavan@yasar.edu.tr

Abstract: In many security protocols, implementation of cryptographic algorithms can be thought as both software and hardware manner. The security of cryptographic algorithms is threatened by attacking cryptographic implementations. In public cryptography, research on side channel attacks started with works of (Kocher et al., 2004, Kocher, 1996, Kocher et al., 1999). Side channel attacks is a passive attack which basically depend on collecting information through physically observable phenomenon's created by the execution of computing tasks in microelectronic devices and logical gates in order to break cryptographic algorithms such as timing information proposed in (Kocher, 1996, Jean-Fran et al., 2000), power consumption studied in (Kocher et al., 1999) and electromagnetic radiation in (Agrawal et al., 2003).This paper serves as an introduction to basic level of side channel attacks, method and techniques used in these attacks and many other countermeasures such as Simple Power Analysis, Differential Power Analysis; Data bit Differential Power Analysis against such attacks.

Keywords: side-channel attacks, countermeasures against side channel attacks

1. Introduction

Since the security has been the main concern in computing systems, much effort is needed to deal with the weaknesses of security protocols and implementations of security algorithms. Cryptographic algorithms, including symmetric ciphers, public-key ciphers, and hash functions, form a set of primitives that can be used as building blocks to construct security mechanisms that target specific objectives(Schneier, November 1996). In general, security protocols deal with what functions are performed during the encryption and decryption. On the other hand, to maintain the security of the computing systems, cryptanalysis of hardware as a part of system should be considered and weaknesses should be defined.

Cryptographic protocols implemented on the hardware include lots of electronic components in order to handle data operations. Cryptanalysis of software part can be made over mathematical abstractions. On the other hand, since a cryptographic algorithm runs on a particular electronic devices cryptanalysis can be done by using physical quantities such as temperature, power consumption, elapsed time during the computation, electromagnetic radiation etc. The main purpose of side channel attacks is extracting the secret key of cryptosystem via gathering physical information's, namely side channel information's, in order to break cryptosystem. Various measurements are taken from different places during the execution of a cryptographic algorithm and these values are used directly for extracting secret key of system.

Having considered classification of side-channel attacks, two main topics are suggested.

1. Invasive and non-invasive: Invasive attack is the strongest type of side-channel attack that is basically a way of de-packing cryptographic device in order to get direct access to components. Connecting a wire on a data bus and being able to capture the data transfer is a simple example of invasive attack. A non-invasive attack is a way to exploiting external available information. A typical example of such an attack is timing analysis which is measuring the time consumed by a device to execute an operation and correlating this with the computation performed by the device in order to deduce the value of the secret keys(Feng, 2005).

2. Passive and active: As the name suggests, passive attacks are done by observing physical quantities of cryptographic system on behalf of exposing secret key. Yet active attacks make the system operate unusually by manipulating incoming and outgoing data of the system.

Bearing in mind the previous points , passive non-invasive attacks are generally referred as side channel attacks that is determining secret key of a system by measuring physical quantities such as power consumption, electromagnetic radiation etc. .

1.1 Source of the leakages

Since microelectronic devices are used for accomplishing computing tasks, some leakages like dissipation of heat occur and these leakages carry lots of information about the computation. By capturing these leakages, it is possible to exploit the whole system by malicious adversaries.

Most of logic circuits have been implemented over CMOS gates which are used to create integrated circuits. Since CMOS gates are the brain of the digital circuits, side channel information can be captured from CMOS during the computation. Emanating side channel information's from CMOS devices distinguished by distinct levels; power consumption in CMOS devices and EM radiation in CMOS devices. Besides, these levels can be examined as dynamic and static.

1.2 Power consumption in CMOS devices:

In CMOS devices, power goes to three main areas.

Dynamic power dissipation: is the amount of power needed in order to perform tasks in a given circuit namely power that need to charge and discharge electronic nodes (capacitors, transistors etc.). Dynamic power consumption connects internal data with externally observable power therefore can be used to get side-channel information. Dynamic power dissipation can be calculated in following way:

$$P_{dynamic} = C_L x V_{dd}^2 x P_{(0-1)} x f$$

C_L Represents capacitance, V_{dd} is the voltage of the power supply and $P_{(0-1)}$ is the probability of changing 0 states to 1. Dynamic power dissipation occurred when switching states. The highest peak of CMOS device occurs when charging the capacitance.

Short Circuit Power Dissipation: Input signals have nonzero rise and fall times; hence, a direct current path exists between V_{dd} and GND for a short period of time during input switching, in which case the PMOS and the NMOS devices are simultaneously conducting(Chen, 1997).

Leakage power: In contrast to dynamic power consumption, leakage of current is not a function of switching frequency namely, leakage of current always occurs.

1.3 EM radiation in CMOS devices

Like in power consumption, electromagnetic radiation is also data dependent. In theoretical view, EM leakages expressed with Biot-Savart law:

$$dB = \frac{\mu I dl x r}{4\pi r^2}$$

Where μ is magnetic permeability of free space, dl infinitesimal length of conductor carrying electric current I , r is the unit vector. This general formula gives an idea for EM radiation in CMOS devices and emphasizes two important points(Standaert, 2009);

- The field is data-dependent due to the dependence on the current intensity.
- The field orientation depends on the current direction.

Due to the fact that radiation is data-dependent, it causes side-channel information leakages. EM attacks first represented by (Quisquater and Samyde, 2001)later on by placing coils near the chip, side-channel information is exploited by (Quisquater and Samyde, 2001)and (Gandolfi et al., 2001)

2. Types of attacks

2.1 Power analysis attacks

Power analysis attack is a passive non-invasive attack that reveals a secret key of the system by analyzing power consumption. Power consumption of a cryptographic device depends on processing data and

performing operation thus instantaneous power analysis can be done and secret key can be constructed in that way. Power analysis attacks have been demonstrated to be very powerful attacks for most straightforward implementations of symmetric and public key ciphers proposed in(Mayer-Sommer, 2000, Messerges et al., 2002, Messerges et al., 1999, Novak, 2002, Walter, 2001).In 1998, Kocher's result of power attack analysis research is presented in New York Times. In this research, Kocher claimed it is possible to reveal the secret key completely from a power trace of a single operation. As a result of this research, Kocher realized that power analysis is a serious threat in cryptographic protocols. Based on Kocher's research simple and differential power analysis (SPA and DPA), first introduced by(Kocher et al., 1999).

Depending on cryptographic operations, the amount of current can easily be captured from a power source. Using an oscilloscope, the voltage difference can be calculated and using Ohm's law current consumption on system can be subtracted(see Figure 1). Power dissipation is proportional to the flow of current in constant voltage devices thus power analysis attacks can be made through measuring the current consumption.

Public-key cryptography is more generally based on exponentiation algorithms that is given group G , $M \in G$ and $d \in \mathbb{Z}$ find $S = M^d mod N$. The basic algorithm for calculating exponentiation is binary method as shown in Algorithm 1.

```
Input: M ∈ G, d =(d_n d_{n-1}..d_1 d_0)_2
Output: S = M^d mod N

1. S = M
2. for i from 1 to n-1
do
3. S = S * S( mod N)
4. if d[i]= 1 then
5.     S = S * M (mod N)
6. return S
```

Algorithm 1: Left-to-right binary exponentiation

For instance in a binary exponentiation algorithm the squaring operations could have a different power output to the general multiplication operations. Thus, it is possible to recover the key by looking the order of bits. For example if we notice the sequence of square and multiply operations was SMSSMSMSS then, assuming the binary method of exponentiation was used, we know that the bits of the exponent are given by 101101011.

2.1.1 Simple power analysis attacks

The key aspect of simple power analysis is making guesses from power trace in which instruction executed and values of input and output data's.

To illustrate the simple power analysis attack, implementation of RSA algorithm used as an example. In 1977, RSA, a public key cryptosystem, developed by Ronald Rivest, Adi Shamir, and Leonard Adleman. RSA algorithm works by taking two large primes let's say p and q, and computes their product n as $n = p.q$ (called modulus). After that, it chooses a random e and must be relatively prime to $(p-1)(q-1)$. Later, it finds another number d such that $(ed-1)$ is divisible by $(p-1)(q-1)$. Numbers e and d is called public-key and private key exponents, respectively. Public-key pair is (n, e) and private key pair is (n, d). The difficulty of breaking RSA algorithm is mainly depends on factorization problem of the number in the case of factoring n to p and q. Algorithm 2 shows the summary of RSA.

```
1. Choose two very large random prime integers:
        p and q
2. Compute n and φ(n):
        n = pq and φ(n) = (p-1)(q-1)
3. Choose an integer e, 1 < e < φ(n) such that:
        gcd(e, φ(n)) = 1
4. Compute d, 1 < d < φ(n) such that:
        ed ≡ 1 (mod φ(n))
5.   The public key is (n, e) and the private key is (n, d)
```

Algorithm 2: RSA algorithm

In Luca Giancane's illustration "An SPA trace of an RSA algorithm"(Giancane, 2012)it is shown that small segment of power trace of RSA algorithm on a smart card. Nine vertical lines show that squaring or multiplication operation is started. In the beginning of algorithm, values are loaded to registers. Multiplication operation is more costly from other operations since it requires additional registers. Thus, in the multiplication operation, width of leading vertical lines increases. Consequently, squaring operations are represented as narrow vertical lines and multiplications are represented as narrow vertical lines followed by wider vertical lines. Finally, five key bits can easily be extracted from the power analysis as 00111.

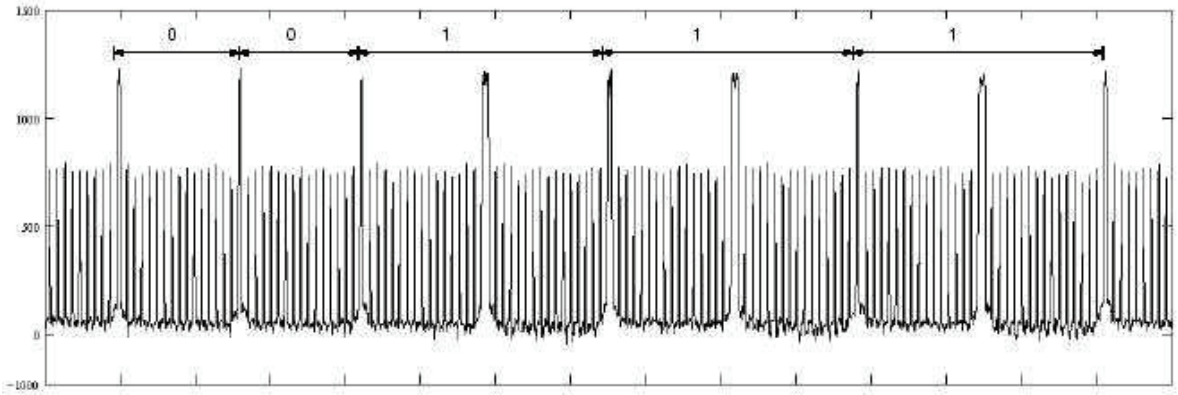

Figure 1: Luca Giancane, "An SPA trace of an RSA algorithm"

Observing power trace of algorithms, an attacker can get more details about algorithms and easily captures specific operation being executed. As a consequence, an attacker would simply produce a secret key of a system by capturing power trace of an operation.

2.1.2 Differential power analysis attacks

On the contrary with SPA, large number of power traces needed to accomplish DPA attacks. Depending on Kocher's research, DPA is the most threatening attack in contrast to other attacks. The main idea behind the differential power attack is that during an encryption or decryption operation of different data values or blocks, a large number of power traces recorded on behalf of revealing secret key. The main advantage of DPA is that, it is generally required to know cryptographic algorithm that is used in cryptographic device. Another difference of DPA from SPA is that the shapes of power traces along time axis is not so important in DPA while an attacker tries to find patterns in SPA. In DPA an attacker tries to find correlation between power consumption at fixed moment and processed data. To sum up, DPA attacks exploit the data dependency of the power consumption of cryptographic devices. They use a large number of power traces to analyze the power consumption at a fixed moment of time as a function of the processed data (Mangard et al., 2007).

When the internal state of device changes, the power consumption covaries regarding the change. For instance, consider the following two state changes: 01010001 to 10101110 and 01010001 to 11110001. In the first example, it requires eight states of changes and second example only requires two states changes in corresponding transistors. The number of change differences highly effects power consumption levels in the cryptographic device. The main idea behind DPA is that finding correlations between the different runs for guessing key by making observations of power trace over a number of runs of the cryptographic algorithm.

After that, if the key guess is correct then one will obtain a high correlation, if it is not then they will not be correlated.

In 1999, (Kocher et al., 1999)were introduced the SPA and DPA attacks. Kocher conducted practical analysis attack by implementing DES in hardware. First implementation of power analysis attack on elliptic curves is done by (Jean-S et al., 1999)and introduced SPA-resistant method for point multiplication. The experimental results of power analysis attacks on smart cards explored (Akkar et al., 2000, Messerges et al., 2002).

2.2 Timing analysis attacks

Running time of cryptographic algorithms may give information about key and secret parameter to attacker. Timing attacks is a side-channel attacks that an attacker observing the time taken by cryptographic algorithms in order to break cryptosystem. Difference between timing attacks and power analysis attacks is that, timing attacks can also be applied to network based cryptosystems therefore it is not restricted to cryptographic tokens. The basic assumptions of timing analysis are represented by(Feng, 2005):

- The run time of a cryptographic operation depends to some extent on the key. With present hardware this is likely to be the case, but note that there are various efficient hardware based proposals to make the timing attack less feasible through 'noise injection' .
- A sufficiently large number of encryptions can be carried out, during which the key does not change. A challenge response protocol is ideal for timing attacks.
- Time can be measured with known error. The smaller the error, the fewer time measurements are required.

This idea was first introduced by (Kocher, 1996), and then implemented against an RSA implementation using the Montgomery algorithm(Akkar et al., 2000). In the timing attacks, an attacker measures running time of each corresponding messages which is processed by device.

In (Koeune, 2002)some reports about timing attacks is proposed: In (Brumley and Boneh, 2003)(Schindler et al., 2001) the improved version of timing attack represented that is basically makes possible to recover a 512-bit key using between 5.000 and 10.000 timing measurements . Also, Schindler represent that it is possible to break a 1024 key approximately 370 time measurements in the implementation of CRT in (Schindler, 2000). Also, block ciphers suffers from timing attacks. Timing attack to RC5 is reported by(Gutmann, 1996) and also in (Schindler et al., 2001)about 4.000 time measurement it is possible recover AES key.

2.3 Electromagnetic analysis attacks

Moving electric charges produces electromagnetic field. The currents going through a processor can characterize it according to its spectral signature. Electromagnetic attacks are first introduced by (Quisquater and Samyde, 2001), and further developed in (Gandolfi et al., 2001). In this study, the attack is conducted by exploiting the side channel by placing coils in the neighborhood of the chip and analyzing the measured electromagnetic field(Koeune, 2002).

As in the analysis of power attacks, electro-magnetic analysis can be made in two ways; simple electro-magnetic analysis and differential electro-magnetic analysis, SEMA and DEMA namely. When cryptographic device is processing computations it leaks EM emanations each represents different information about operations. EM emanations can be categorized in two levels: direct emanations and unintentional emanations. Direct emanations arise from the flow of the current. On the other hand unintentional emanations arise from components which are close to each other. In 1950, it was shown that it is possible to measure electromagnetic radiation from a smart card by (Quisquater and Samyde, 2001)and(Gandolfi et al., 2001). Furthermore, the work of (Agrawal et al., 2003)showed that EM attacks can be used to attack where the power analysis of the cryptographic device is not available and also EM can even be used to break power analysis countermeasures.

To conclude, an attacker can get plenty of side channel information with exploring EM emanations. Also, it is easy to make table of instructions with using power/EM traces and correlation techniques which makes breaking the key easier.

2.4 Fault attacks

Most of time, it is assumed that cryptographic devices operate in a reliable way. Since cryptographic devices include electronic components, corresponding hardware faults can effect operations which means security of these devices somewhat depend on the reliability of implementation of the algorithm. In cryptographic manner, faulty computations can take an advantage to discover a secret key. A fault attack plays an important role on behalf of cryptanalysis. The idea behind fault attacks as means of cryptanalysis is that tampering cryptographic device in order to execute erroneous operations and hoping that this erroneous behavior does not leak information about secret parameters.

On secret key cryptosystems, differential fault attacks are introduced by (Biham and Shamir, 1997). By analyzing between 50 and 200 cipher texts, they recovered full DES key after performing differential fault attack on DES. In (Boneh et al., 1997), the authors showed how to factor RSA modulus by using one faulty and one correct signature. Also in(Koeune, 2002), papers about fault attacks referred to readers (Anderson, 2001, Anderson and Kuhn, 1996, Anderson and Kuhn, 1998, Gutmann, 1996, Gutmann, 2001, Maher, 1997).

3. Countermeasures

Both in hardware and software, there are many strategies exist to prevent side-channel attacks. Possible solutions and their efficiency are described in this section. Generally, software based countermeasures include inserting dummy instructions, balancing Hamming Weight of internal data, bit splitting. On the hardware level countermeasures include randomization of instruction set execution, clock randomization, power consumption randomization. A large variety of solutions exist in order to prevent side-channel attacks and no single one is enough for the perfect security.

In the work of (Borst, 2001)and(Mangard, 2004), there exists a nice survey for countermeasures. Time randomization is proposed in(Chari et al., 1999, Vincent Rijmen, 1999, Ha and Moon, 2003, Izu and Takagi, 2002a, Liardet and Smart, 2001, Oswald and Aigner, 2001, Benini et al., 2003),permuting the execution is in(Goubin and Patarin, 1999) and masking technique is in (Jean-S et al., 2000, Golic, Golic and Tymen, 2003, Goubin, 2001, Messerges, 2001, Trichina et al., 2003, Akkar and Giraud, 2001)(Izu and Takagi, 2002b)which are the main software countermeasures against power attacks. Works against power analysis attacks proposed by power signal filtering in (Jean-S et al., 2000, Shamir, 2000)and in(Jacques J.A. Fournier, 2003, Moore et al., 2003), asynchronous circuits used.

At the different abstraction level, a few suggestions in order to prevent side-channel attacks proposed in (Verbauwhede, 2010):

- In order to improve the resistance of a physical devices, shields, conforming glues cited from (Anderson and Kuhn, 1996)_ENREF_31, physically unclonable functions cited from (Tuyls et al., 2006), detectors, detachable power supplies cited from (Shamir, 2000) can be used.
- On the technological level, dynamic and differential logic styles (as an alternative to CMOS) have been proposed in various shapes (e.g., (Tiri, 2002)) to decrease the data dependencies of the power consumption.
- On the algorithmic level, time randomization is cited from (May et al., 2001), encryption of the buses cited from (Brier et al., 2001),hiding (i.e., making the leakage constant), or masking (i.e., making the leakage dependent of some random value, e.g. in(Goubin and Patarin, 1999)) are the usual countermeasures.
- At all the previous levels, noise addition is the generic solution to decrease the amount of information in the side-channel leakages.
- Countermeasures also exist at the protocol level, e.g., based on key updates.

Also, according to(Feng, 2005) general strategies of countermeasures both hardware and software is proposed and cited from (Goubin and Patarin, 1999):

- De-correlate the output traces on individual runs (e.g., by introducing random timing shifts and wait states, inserting dummy instructions, randomization of the execution of operations etc.);
- Replace critical assembler instructions with ones whose “consumption signature” is hard to analyze, or re-engineer the critical circuitry which performs arithmetic operations or memory transfers;

- Make algorithmic changes to the cryptographic primitives so that attacks are provably inefficient on the obtained implementation, e.g., masking data and key with random mask generated at each run.

Countermeasures against side-channel attacks are summarized by (Feng, 2005): "Software-based countermeasures include introducing dummy instructions, randomization of the instruction execution sequence, balancing Hamming weights of the internal data, and bit splitting. On the hardware level, the countermeasures usually include clock randomization, power consumption randomization or compensation, randomization of instruction set execution and/or register usage." (p. 23)

4. Conclusion

To sum up foregoing, this paper reviews the general aspect of side-channel attacks and weaknesses of the systems. In the countermeasures section general solutions are proposed by various sources to indicate making side-channel attacks harder and limit the possible threats. Due to the fact that side-channel attacks based on natural properties of physical phenomenon's and leakages, there exist some way to get side-information like timing attacks; power analysis, em analysis and inserting fault attacks. Power consumption and em radiation of systems directly related with the usage of data. By applying statistical tools to these leakage models, it is possible to reveal the key of systems. To conclude, this paper presents a general perspective for causes of side-channel attacks and side-channel attacks and summarizes the literature on side channel attacks.

References

Agrawal, D., Archambeault, B., Rao, J. R. & Rohatgi, P. 2003. The Em Side-Channel(S). Revised Papers From The 4th International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Akkar, M.-L., #233, Bevan, G., Dischamp, P. & Moyart, D. 2000. Power Analysis, What Is Now Possible. Proceedings Of The 6th International Conference On The Theory And Application Of Cryptology And Information Security: Advances In Cryptology. Springer-Verlag.

Akkar, M.-L. & Giraud, C. 2001. An Implementation Of Des And Aes, Secure Against Some Attacks. Proceedings Of The Third International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Anderson, R. & Kuhn, M. 1996. Tamper Resistance: A Cautionary Note. Proceedings Of The 2nd Conference On Proceedings Of The Second Usenix Workshop On Electronic Commerce - Volume 2. Oakland, California: Usenix Association.

Anderson, R. J. 2001. Security Engineering: A Guide To Building Dependable Distributed Systems, John Wiley \&Amp; Sons, Inc.

Anderson, R. J. & Kuhn, M. G. 1998. Low Cost Attacks On Tamper Resistant Devices. Proceedings Of The 5th International Workshop On Security Protocols. Springer-Verlag.

Benini, L., Macii, A., Macii, E., Omerbegovic, E., Pro, F. & Poncino, M. 2003. Energy-Aware Design Techniques For Differential Power Analysis Protection. Proceedings Of The 40th Annual Design Automation Conference. Anaheim, Ca, Usa: Acm.

Biham, E. & Shamir, A. 1997. Differential Fault Analysis Of Secret Key Cryptosystems. Proceedings Of The 17th Annual International Cryptology Conference On Advances In Cryptology. Springer-Verlag.

Boneh, D., Demillo, R. A. & Lipton, R. J. 1997. On The Importance Of Checking Cryptographic Protocols For Faults. Proceedings Of The 16th Annual International Conference On Theory And Application Of Cryptographic Techniques. Konstanz, Germany: Springer-Verlag.

Borst, J. 2001. Block Ciphers: Design, Analysis, And Side-Channel Analysis., Katholieke Universiteit Leuven.

Brier, E., Handschuh, H. & Tymen, C. 2001. Fast Primitives For Internal Data Scrambling In Tamper Resistant Hardware. Proceedings Of The Third International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Brumley, D. & Boneh, D. 2003. Remote Timing Attacks Are Practical. Proceedings Of The 12th Conference On Usenix Security Symposium - Volume 12. Washington, Dc: Usenix Association.

Chari, S., Jutla, C. S., Rao, J. R. & Rohatgi, P. 1999. Towards Sound Approaches To Counteract Power-Analysis Attacks. Proceedings Of The 19th Annual International Cryptology Conference On Advances In Cryptology. Springer-Verlag.

Chen, W.-K. 1997. The Electrical Engineering Handbook, Crc Press.

Feng, Y. Z. A. D. 2005. Side-Channel Attacks: Ten Years After Its Publication And The Impacts On Cryptographic Module Security Testing.

Gandolfi, K., Mourtel, C. & Olivier, F. 2001. Electromagnetic Analysis: Concrete Results. Proceedings Of The Third International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Giancane, L. 2012. Side Channel Attacks And Countermeasures In The Design Of Secure Ic's Devices For Cryptographic Applications, Lap Lambert Academic Publishing.

Golic, J. D. A New Paradigm For Key-Dependent Reversible Circuits. Springer-Verlag.

Golic, J. D. & Tymen, C. 2003. Multiplicative Masking And Power Analysis Of Aes. Revised Papers From The 4th International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Goubin, L. 2001. A Sound Method For Switching Between Boolean And Arithmetic Masking. Proceedings Of The Third International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Goubin, L. & Patarin, J. 1999. Des And Differential Power Analysis (The "Duplication" Method). Proceedings Of The First International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Gutmann, P. 1996. Secure Deletion Of Data From Magnetic And Solid-State Memory. Proceedings Of The 6th Conference On Usenix Security Symposium, Focusing On Applications Of Cryptography - Volume 6. San Jose, California: Usenix Association.

Gutmann, P. 2001. Data Remanence In Semiconductor Devices. Proceedings Of The 10th Conference On Usenix Security Symposium - Volume 10. Washington, D.C.: Usenix Association.

Ha, J.-C. & Moon, S.-J. 2003. Randomized Signed-Scalar Multiplication Of Ecc To Resist Power Attacks. Revised Papers From The 4th International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Izu, T. & Takagi, T. 2002a. A Fast Parallel Elliptic Curve Multiplication Resistant Against Side Channel Attacks. Proceedings Of The 5th International Workshop On Practice And Theory In Public Key Cryptosystems: Public Key Cryptography,Springer-Verlag.

Izu, T. & Takagi, T. 2002b. A Fast Parallel Elliptic Curve Multiplication Resistant Against Side Channel Attacks. Proceedings Of The 5th International Workshop On Practice And Theory In Public Key Cryptosystems: Public Key Cryptography. Springer-Verlag.

Jacques J.A. Fournier, S. M., Huiyun Li, Robert Mullins, And George Taylor 2003. Security Evaluation Of Asynchronous Circuits. Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Jean-Fran, #231, Dhem, O., Fran, #231, Koeune, O., Leroux, P.-A., Mestr, P., #233, Quisquater, J.-J. & Willems, J.-L. 2000. A Practical Implementation Of The Timing Attack. Proceedings Of The The International Conference On Smart Card Research And Applications. Springer-Verlag.

Jean-S, #233 & Coron, B. 1999. Resistance Against Differential Power Analysis For Elliptic Curve Cryptosystems. Proceedings Of The First International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Jean-S, #233, Coron, B. & Goubin, L. 2000. On Boolean And Arithmetic Masking Against Differential Power Analysis. Proceedings Of The Second International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Kocher, P., Lee, R., Mcgraw, G. & Raghunathan, A. 2004. Security As A New Dimension In Embedded System Design. Proceedings Of The 41st Annual Design Automation Conference. San Diego, Ca, Usa: Acm.

Kocher, P. C. 1996. Timing Attacks On Implementations Of Diffie-Hellman, Rsa, Dss, And Other Systems. Proceedings Of The 16th Annual International Cryptology Conference On Advances In Cryptology. Springer-Verlag.

Kocher, P. C., Jaffe, J. & Jun, B. 1999. Differential Power Analysis. Proceedings Of The 19th Annual International Cryptology Conference On Advances In Cryptology. Springer-Verlag.

Koeune, J.-J. Q. A. F. 2002. Side-Channel Attacks: State-Of-The-Art.

Liardet, P.-Y. & Smart, N. P. 2001. Preventing Spa/Dpa In Ecc Systems Using The Jacobi Form. Proceedings Of The Third International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Maher, D. P. 1997. Fault Induction Attacks, Tamper Resistance, And Hostile Reverse Engineering In Perspective. Proceedings Of The First International Conference On Financial Cryptography. Springer-Verlag.

Mangard, S. 2004. Securing Implementations Of Block Ciphers Against Side-Channel Attacks. Institute For Applied Information Processing And Communications(Iaik).

Mangard, S., Oswald, E. & Popp, T. 2007. Power Analysis Attacks: Revealing The Secrets Of Smart Cards (Advances In Information Security), Springer-Verlag New York, Inc.

May, D., Muller, H. L. & Smart, N. P. 2001. Random Register Renaming To Foil Dpa. Proceedings Of The Third International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Mayer-Sommer, R. 2000. Smartly Analyzing The Simplicity And The Power Of Simple Power Analysis On Smartcards. Proceedings Of The Second International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Messerges, T. S. 2001. Securing The Aes Finalists Against Power Analysis Attacks. Proceedings Of The 7th International Workshop On Fast Software Encryption. Springer-Verlag.

Messerges, T. S., Dabbish, E. A. & Sloan, R. H. 1999. Power Analysis Attacks Of Modular Exponentiation In Smartcards. Proceedings Of The First International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Messerges, T. S., Dabbish, E. A. & Sloan, R. H. 2002. Examining Smart-Card Security Under The Threat Of Power Analysis Attacks. Ieee Trans. Comput., 51, 541-552.

Moore, S. W., Anderson, R. J., Mullins, R. D., Taylor, G. S. & Fournier, J. J. A. 2003. Balanced Self-Checking Asynchronous Logic For Smart Card Applications.

Novak, R. 2002. Spa-Based Adaptive Chosen-Ciphertext Attack On Rsa Implementation. Proceedings Of The 5th International Workshop On Practice And Theory In Public Key Cryptosystems: Public Key Cryptography. Springer-Verlag.

Oswald, E. & Aigner, M. 2001. Randomized Addition-Subtraction Chains As A Countermeasure Against Power Attacks. Proceedings Of The Third International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Quisquater, J.-J. & Samyde, D. 2001. Electromagnetic Analysis (Ema): Measures And Counter-Measures For Smart Cards. Proceedings Of The International Conference On Research In Smart Cards: Smart Card Programming And Security. Springer-Verlag.

Schindler, W. 2000. A Timing Attack Against Rsa With The Chinese Remainder Theorem. Proceedings Of The Second International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Schindler, W., Fran, #231, Koeune, O. & Quisquater, J.-J. 2001. Improving Divide And Conquer Attacks Against Cryptosystems By Better Error Detection / Correction Strategies. Proceedings Of The 8th Ima International Conference On Cryptography And Coding. Springer-Verlag.

Schneier, B. November 1996. Applied Cryptography: Protocols, Algorithms And Source Code In C, John Wiley And Sons.

Shamir, A. 2000. Protecting Smart Cards From Passive Power Analysis With Detached Power Supplies. Proceedings Of The Second International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Standaert, F.-X. 2009. Introduction To Side-Channel Attacks. In: Verbauwhede, I. M. R. (Ed.) Secure Integrated Circuits And Systems Springer.

Tiri, K., Akmal, M. ; Verbauwhede, I. 2002. A Dynamic And Differential Cmos Logic With Signal Independent Power Consumption To Withstand Differential Power Analysis On Smart Cards. Solid-State Circuits Conference, 2002. Esscirc 2002. Proceedings Of The 28th European. Florence, Italy.

Trichina, E., Seta, D. D. & Germani, L. 2003. Simplified Adaptive Multiplicative Masking For Aes. Revised Papers From The 4th International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Tuyls, P., Schrijen, G.-J., Boris, #352, Kori, #263, Geloven, J. V., Verhaegh, N. & Wolters, R. 2006. Read-Proof Hardware From Protective Coatings. Proceedings Of The 8th International Conference On Cryptographic Hardware And Embedded Systems. Yokohama, Japan: Springer-Verlag.

Verbauwhede, I. M. R. 2010. Integrated Circuits And Systems, Springer.

Vincent Rijmen, J. D. 1999. Resistance Against Implementation Attacks: A Comparative Study Of The Aes Proposals. In Proc.The Second Advanced Encryption Standard(Aes) Candidate Conference.

Walter, C. D. 2001. Sliding Windows Succumbs To Big Mac Attack. Proceedings Of The Third International Workshop On Cryptographic Hardware And Embedded Systems. Springer-Verlag.

Models for the Forensic Monitoring of Cloud Virtual Machines

Dirk Ras and Hein Venter
Information and Computer Security Architectures Research Group, Department of Computer Science, University of Pretoria, South Africa
dras@cs.up.ac.za
hventer@cs.up.ac.za

Abstract: With the prevalence of cloud computing and the "as a Service" delivery paradigm, the need has arisen for the ability to conduct an effective digital forensic investigation on these systems. Linked to this is the concept of forensic readiness, which when implemented correctly, makes it possible to conduct a credible forensic investigation on such large scale systems. This paper proposes five models that will enable forensic monitoring, that with their application will facilitate forensic readiness. These models are designed in a manner that data from guest virtual machines can be captured and stored should the need for forensic investigations arise. The cloud hypervisor plays a pivotal role hosting of the guest virtual machines and thus its native abilities are expanded on to make it possible to capture, catalogue and store information of the guest virtual machines. With the research already done in the field of forensic readiness of cloud computing systems, these models can serve as a possible implementation solution for further research into the field. Five models were chosen as it is the number of places where a forensic monitor can be implemented by a cloud service provider on standard cloud architecture. The proposed models can be implemented on the guest virtual machine operating system, the cloud hypervisor, the communication layer between the cloud operating system and hypervisor, and as a single or multi-tenant forensic virtual machine. Because of the cost associated with implementing new hardware, the proposed models are all software based and can be implemented with existing cloud infrastructures without the need to change hardware configurations. For the purposes of this paper a forensic investigation is of a corporate nature where fraud or disaster recovery is the primary motivation. With the case of government security agency involvement a different motivation would apply which is not relevant to this paper.

Keywords: digital forensics, computer forensics, hypervisor, cloud computing, forensic monitoring

1. Introduction

Since the early 80s, computer forensic science has grown from analysis of low capacity hard drives, to today, where massive cloud clusters run into peta-bytes of information. With this, the need has grown to be able to perform computer forensics accurately, affordably and in a timely manner. To achieve this, forensic readiness has been brought to the forefront. Forensic readiness can be defined as when an organisation has implemented certain procedures that would facilitate the digital forensic process. It can also assist in the prediction of unauthorized action that can be disruptive to normal operations (Pangalos et al. 2010). Forensic investigations are likely to be time consuming and expensive, thus companies use them only when all other options of solving a problem has been exhausted. Since with adequate planning and strategy, the implementation of forensic readiness is achievable, the problems of high cost and time can be mitigated, as stated by (Pangalos et al. 2010), (Kruse & Heiser 2002) and (Ruan et al. 2011).

The added complexity of multi-tenant cloud computing systems exasperates the problems that are already faced by forensic investigators and cloud service providers. It would not be feasible to shut down an entire cloud cluster to perform a forensic investigation, assuming the investigation takes place in a normal corporate environment. The impact in terms of financial loss and service down time could be disastrous to the service provider. Also with the different service models available from cloud providers, there needs to be agility in the methods of forensic techniques. In light of these factors, strategies to simplify and speed-up the forensic process need to be considered (Kruse & Heiser 2002).

In this paper, five models are proposed with the purpose of monitoring information of a virtual machine (VM) deployed in a cloud hosting environment. These models offer a possible solution to the problems mentioned above. Five models were chosen as each model can be implemented as a software solution on the standard cloud architecture. All the models were designed in a way that they can be created by a single entity among the role players, be it the cloud service provider, the hypervisor creator or the operating system (OS) creator. All the models contain a forensic monitor that facilitates forensic capturing of data in a virtual machine. The models are developed with incident handling in mind and criminal investigations will remain outside the scope of this research. A sixth model is also possible where the forensic model is hardware based. This model falls outside the scope of this paper.

This paper is structured as follows: section 2 contains background information on the various components used in the models, section 3 contains the proposed forensic models and their descriptions, section 4 contains a discussion of the models, section 5 contains a summary table and section 6 contains possible future work to be done to realize the proposed models. This paper is an initial research piece, aimed at getting a theoretical base to later implement and evaluate the described models. As such no recommendation can as yet be made as to which model is best.

2. Background

This section gives the background information regarding the concepts underpinning the forensic models. The core concepts are that of cloud computing, cloud data delivery paradigms and digital forensics.

2.1 Cloud computing

Cloud computing is a model that can be defined as a method to gain on-demand access to a shared pool of resources with minimal effort from the user and minimal management input from the service provider (Mell & Grance 2011) (Ras & Olivier 2012). The core technologies required for cloud computing to be realised is clustering and virtualization of IT infrastructure (Hess & Newman 2009). The control mechanism to handle the resource allocation of the virtualized cluster is the hypervisor (Van Staden & Venter 2011).

Clustering allows multiple computers to form the resource pool needed to enable cloud computing (Manoel et al. 2002) (Ras & Olivier 2012) (Moreno-Vozmediano et al. 2009). This is done be interconnecting a large number of nodes to perform this funtion (Manoel et al. 2002). These resources are then abstracted for use by virtualization (Lillard 2010). The virtualiszed resources can then be provisioned for use in the form of VMs (Lagar-Cavilla et al. 2011) (Ruan et al. 2011). The hypervisor controls the provisioning and resource allocation process (Szefer & Lee 2012). The VMs deployed via the hypervisor, are the guest VMs (Szefer & Lee 2012) (Ruan et al. 2011).

2.2 Cloud data delivery paradigms

In cloud computing there are several data delivery paradigms which are fundamental to the delivery of services. These paradigms are Software-, Platform- and Infrastructure as a Service (Mell & Grance 2011) (Voorsluys et al. 2011). Of the service delivery paradigms, Infrastructure as a Service is the most basic and offers virtualised resources that can be configured in the same manner as a physical server (Nurmi et al. 2009) (Sotomayor et al. 2009). Platform as a Service provides the user with a computing platform and solution stack which said user can control and configure (Chang et al. 2010) (Mell & Grance 2011). Finally, Software as a Service provides a software solution to a user which is accessed via some client application, like a web browser (Nurmi et al. 2009) (Hayes 2008).

2.3 Digital forensics

Digital forensics is a recognized scientific and forensic process used in digital forensic investigations (Ballou et al. 2001). For the purposes in this paper and since a formal definition has not yet been agreed upon in the literature, the terms digital forensics and computer forensics can be used interchangeably (Andrew 2007). This paper will focus on the concept of forensic readiness of systems hosted in a cloud environment.

2.3.1 Forensic readiness

As cloud computing is a business model similar to normal hosting, where individual pieces of computer hardware are rented to customers, it is possible to collect evidence in the form of log files, network traffic and system transactions for the use in forensic investigations. Forensic readiness make it possible for organisations to respond should the need arise to perform a forensic investigation (Pangalos et al. 2010) (Rowlingson 2004). Readiness can be used as a method to pre-empt the process of forensic data collection as cost and time saving strategy.

2.3.2 Forensic monitoring

Forensic monitoring can be used alleviate the problem that forensic investigators must sift through large amounts data (Van Staden & Venter 2011). Forensic monitoring tools would capture and store pre-defined

data such as logs and transactions for use in investigations. This pre-processed data would cut down on the time investigators would have to spend during the acquisition phase of an investigation.

2.3.3 Data segregation

Data segregation refers to the separation and isolation of data from different users. It is imperative that the origin of the data be clear. With this, data from each origin source should be kept separate and not mixed as this could lead to cross-contamination between sources. Should cross-contamination occur it could render the value of the forensic data useless (Cohen 2009) (Ruan et al. 2011).

3. Forensic models

In this section the various models for forensic monitoring are described. It will be shown that the models can enable the forensic monitoring of guest VMs. The forensic monitoring of the guest VMs will thus also contribute to the forensic readiness of the cloud system. Each of the models have the same underpinnings in terms of cloud hosting architecture. With this architecture, the discrete hardware nodes are clustered. The cloud operating system runs atop the cluster and provides the hosting platform for the hypervisor. The hypervisor can then host guest VMs enabling the operation of the cloud. This architecture will form the basis of all the proposed models in the following sections. The proposed models will be discussed by the method of implementation, implementation complexity, method of data segregation, resistance to tampering and the relevant service delivery paradigm.

3.1 Operating system embedded forensic monitor

Figure 1 shows the operating system embedded forensic monitor. The forensic monitor forms part of the deployed guest OS. For this method to be feasible, the guest OS must be supplied by the cloud provider. The reason for this is that the user must not be able to access the forensic monitor lest it be tampered with or disabled.

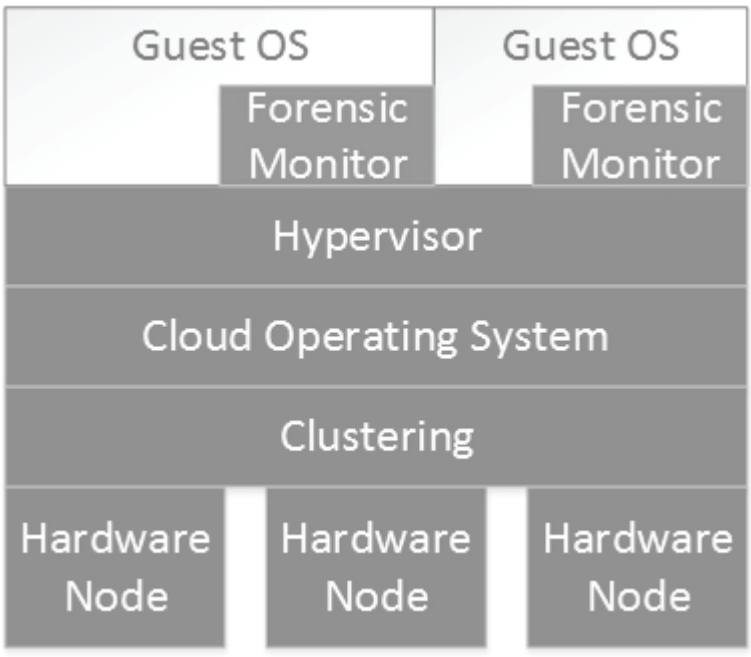

Figure 1: OS embedded forensic monitor

As the forensic monitor is integrated with the guest OS, it gains access to the functions provided by the guest OS. It would thus be possible to capture the relevant forensic data via the guest OS and store it for later use. In terms of data segregation, this model is well suited as the forensic monitor is deployed within each deployed guest OS. Should the forensic data be required, it can be uniquely identified by its originating OS. This model is most suited to the software as a service data delivery paradigm, as the user only consumes the software that is delivered while having no access to the underlying OS. This will hold true only if the OS is secured. The OS might still be vulnerable to attack via exploits such as worms, viruses, etc.

3.2 Hypervisor embedded forensic monitor

Figure 2 depicts the hypervisor embedded model. With this model, the forensic monitor forms part of the hypervisor. As the guest OSs must communicate via the hypervisor to the physical hardware, all traffic can be intercepted. The hypervisor could be engineered in such a manner that the data relevant to forensics is captured and stored. As the monitor is tightly coupled to the hypervisor, the monitor must be written as an integral part of the hypervisor.

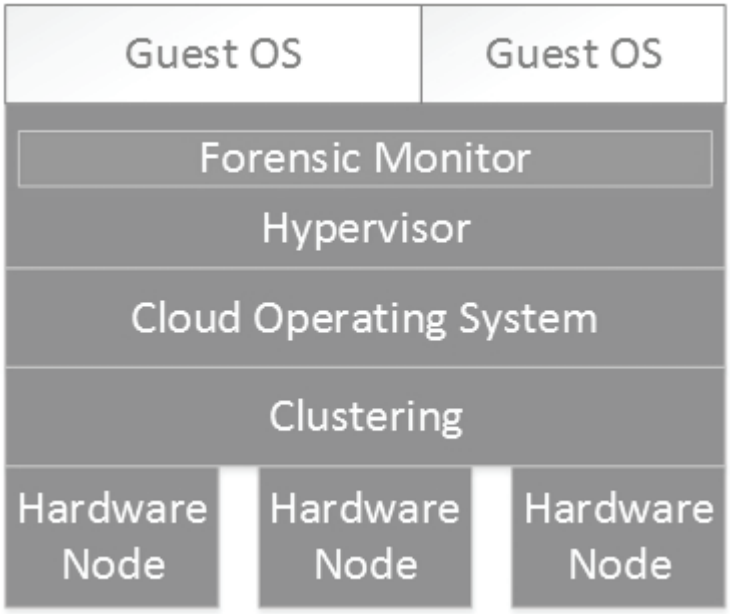

Figure 2: Hypervisor embedded forensic monitor

Data segregation should be taken note of in the design of the forensic monitor. Since the hypervisor is a multi-tenant system, the data from the various guest OSs running on said hypervisor, must be clearly segregated. Since the forensic monitor is integrated into the hypervisor, the user is not limited by which guest OS to use. Unlike the OS embedded monitor model, this model is agnostic to the deployed guest VM making it feasible for the platform as a service data delivery paradigm.

3.3 Communication layer forensic monitor

Figure 3 depicts the communication layer model for forensic monitoring. In this model the forensic monitor is layered between the cloud OS and the hypervisor. With the hypervisor embedded model, the data had to pass from the guest OS to the hardware via the hypervisor and the cloud OS. This model places the forensic monitor as part of the cloud OS, thus forming the communication layer between the cloud OS and the hypervisor. Since all data must pass through the cloud OS to the underlying hardware, the data can be forensically monitored.

With this model, the forensic monitor should form part of the cloud OS to enable the capture of communication between the cloud OS and the hypervisor. This would require that the cloud OS either be designed with the forensic model as part of the system, or that the forensic monitor run as a separate application on the cloud OS.

Data segregation might become a problem with this model. The data should be clearly differentiated by its originating guest OS so that it can be catalogued correctly. Should the data be mixed, it could call into question its forensic usefulness.

In the case that the forensic monitor is tightly coupled with the cloud OS, this model would be suited to the infrastructure as a service data delivery paradigm. This is because the forensic monitor forms part of the cloud OS and cannot be removed. However, in the case of the forensic monitor being a separate application running on the cloud OS, the platform as a service data delivery paradigm would be used. This is due to the fact that as a separate application, the forensic monitor can much more easily be tampered with or disabled.

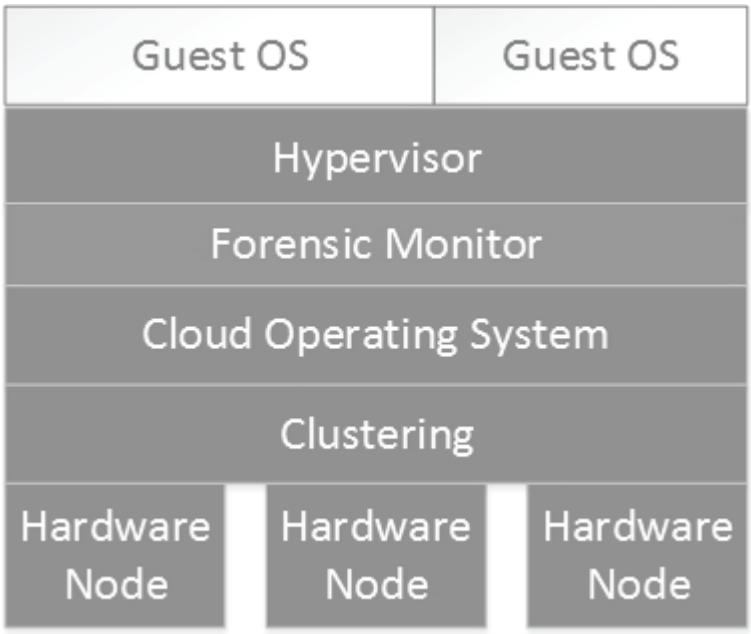

Figure 3: Communication layer forensic monitor

3.4 Single tenant forensic virtual machine

Figure 4 depicts the single tenant mode. This model has the monitoring done by a forensic VM running on the hypervisor. The forensic VM serves as the host to the guest VM and the forensic monitor. This creates a nested set of VMs running on the hypervisor. For this model, each guest VM is hosted by a separate forensic VM.

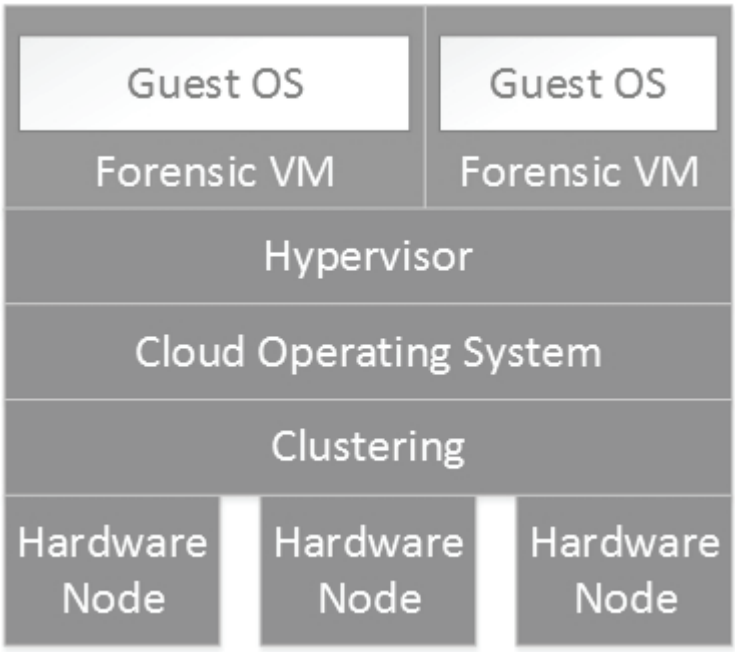

Figure 4: Single tenant forensic virtual machine

As the guest OS is contained within the forensic VM, it is completely isolated from the underlying architecture, making tampering with the forensic monitor very difficult. The forensic monitor gains access to the data of the guest VM, since the data must pass through the forensic VM in order to reach the hardware layers.

In terms of implementation, since the hypervisor is agnostic to the guest OS it hosts, the forensic VM can be run atop the hypervisor. Part of the forensic VM would then contain a hypervisor to enable the hosting of the guest VM. The forensic VM OS must be written in a way that it can fulfil its function of capturing the forensic

data. A possible problem with this model could be that since the VMs are nested, the computational overhead required to run the forensic VM, might cause degradation in performance.

As the forensic VM contains an OS running the forensic monitor, it would allow all the functions associated with such an OS. Thus, captured data can be catalogued and stored for easy access should the need for forensics arise. With this, data can be transported to another system for safe keeping. As each guest VM runs in a separate forensic VM, data segregation is not an issue. All data is automatically kept separate between the forensic VMs, since each forensic VM runs only a single instance guest VM.

As any OS can be deployed in the forensic VM, the user is not constrained in choosing a desired guest OS. Thus, this model would be well suited to the platform as a service data delivery paradigm, as the user can freely select the appropriate guest OS.

3.5 Multi-tenant forensic virtual machine

Figure 5 depicts the multi-tenant model. As with the single tenant model, the multi-tenant VM model has a nested set of VMs running on the hypervisor. In this case however, the forensic VM hosts multiple guest VMs.

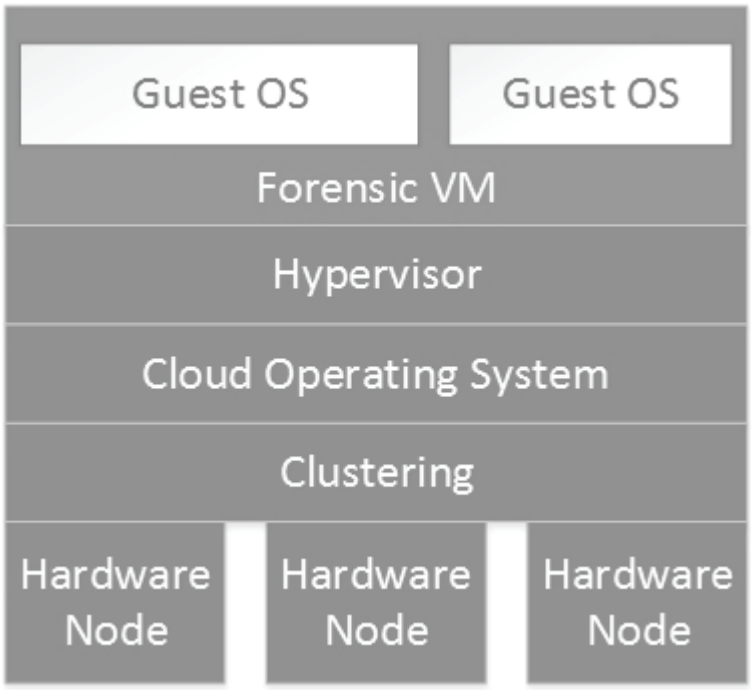

Figure 5: Multi-tenant forensic virtual machine

These hosted guest VMs are isolated form the underlying architecture by the forensic VM. The forensic VM captures the relevant forensic data from the hosted guest VMs and stores it for later use. What is unique about this model is that the guest VMs can directly interact with one another creating a virtual ecosystem. This entire ecosystem can then in turn be monitored by the forensic VM as it would facilitate the communications between the guest VMs.

Data segregation with this model can be an issue, since the forensic VM is a multi-tenant system, the data must pass through the forensic VM to the underlying architecture. Thus, some method must be considered to differentiate between the data from the different guest VMs.

The user can decide on the OS which is to be deployed in the forensic VM as the forensic VM is agnostic to what it hosts. Thus the multi-tenant model is well suited to the platform as a service data delivery paradigm.

4. Discussion

In section 3 the various models were discussed in terms of their construction, implementation complexity, level of data segregation, tamper resistance and service delivery paradigm. This section will compare the models using the above mentioned criteria.

4.1 Implementation

Each of the models differ slightly in terms of implementation, which is to be expected, as this is the primary differentiating factor. The OS Embedded model has the forensic monitor as part of the guest OS and runs as an application on the guest OS.

The Hypervisor Embedded model has the monitor programmed into the hypervisor. It forms part of the construction of the hypervisor and its operation.

The Communication Layer model can be implemented as either part of the cloud OS as an application or service, or it can be embedded into the hardware of the cloud system.

The Single- and Multi-tenant VM models have similar implementations in that both run as virtual machines on the hypervisor. A requirement would be that the forensic virtual machines must support virtualization in order to host a guest OS. A difference would be that the Single tenant model would host only a single guest virtual machine and the Multi-tenant model would host more than one guest OS.

4.2 Implementation complexity

One of the goals of the forensic monitoring models is to be implemented as quickly as possible. It must also be considered that the models will in all likelihood have to be applied to existing systems. Thus, implementation complexity takes into account the speed at which a model can be deployed and the effort required in the deployment. From the complexity point of view, it would be more complex to implement a hardware component than it would be to install a software application

The OS Embedded model would have some complexity in its implementation. As the monitor must form part of the OS it must be added in such a way that it cannot easily be disabled or tampered with. This will require some OS knowledge and design skills.

The Hypervisor Embedded model would be quite complex to implement as it should form an integral part of the hypervisor. A hypervisor developer would be best suited to add the forensic monitor to their product as it can be designed from the ground up to be integrated into the hypervisor software.

The Communication layer model would also be quite complex to implement. For the case where the monitor forms part of the cloud OS, the cloud OS should be designed from the ground up to have the forensic monitor included. For the hardware case, the monitor would have to form part of the communication protocol. In both cases the best solution to the implementation problem would be for the developers of the cloud OS and the hardware to embed the forensic monitor into their respective systems.

For both the Single- and Multi-tenant virtual machine models the implementation would be quite simple. In both cases the forensic VM can merely be run on the hypervisor and the guest VM deployed within it.

4.3 Data segregation

The data captured from the guest VMs in multi-tenant systems must be kept separated as to avoid cross contamination.

The OS Embedded model would offer a high level of data segregation. As each guest OS would contain a forensic monitor, the monitor can add information that uniquely identifies the information of its assigned guest OS.

Depending on the implementation of the Hypervisor Embedded model, data segregation can be low to high. Should the hypervisor only act as a communication conduit with no intelligence as to the data origin or

destination, the data would be mixed. On the other hand, should the hypervisor contain some intelligence as to the origin and destination of the data, it could easily segregate the data of the different guest VMs.

Cognisance as to data segregation would have to be taken with the Communication layer model. If the monitor is embedded in the cloud OS, the cloud OS should have some way of differentiating the data from the hypervisor. In the case of hardware implementation, the data segregation would be low unless the communications protocol can support the necessary data segregation. Both cases would be implementation dependent.

The Single tenant VM model would offer excellent data segregation. Since each forensic VM contains a single guest VM, all data that flows to and from said guest VM can be monitored and has no chance of being mixed with data form other VMs.

With the Multi-tenant VM model, the data segregation might be an issue. As a forensic VM can contain multiple guest OSs, the data from each OS must be kept track of and then segregated to be forensically useful.

4.4 Tamper resistance

The forensic monitor should serve as both a control mechanism for the cloud service provider and a useful utility that can give the user piece of mind regarding their data. However, systems might be used for malicious purposes, necessitating that the forensic monitor be secure form tampering.

Of all the models the OS Embedded model offers the least amount of tamper resistance. Since the user has access to the OS, it can be changed and exploits can be found to disable, or somehow compromise the forensic monitor.

The Hypervisor embedded, Single- and Multi- tenant VM models all exhibit a high resistance to tampering. In all cases the user does not have access to the underlying architecture and can thus not change it.

Tamper resistance with regards to the Communication Layer model would be highly dependent on the implementation of the model. Should the model run as a separate application it might be possible to access and compromise said application. This can be done by either disabling communication to the forensic application or by feeding the application false information. An embedded approach would have a high tamper resistance. This is because, if the system is designed correctly, there would be fail safe measures, like complete disabling of communications should the monitor fail or be disable. Either way the problem would be detected by standard cloud monitoring tools and alert an administrator.

4.5 Service delivery paradigm

As the models differ in construction, their cloud service delivery paradigms also differ. The OS Embedded model is best suited to the software as a service paradigm. Since the OS is supplied by the cloud service provider, the user only has access to the top layer of the service delivery architecture. Thus, the software as a service paradigm is best suited for this model

The Hypervisor Embedded model allows the user to deploy their own OS onto the hypervisor. As the user is free to choose their desired OS, the service paradigm would be that of platform as a service.

The Communication layer forensic model allows forensic monitoring to take place via the system communication channels. Depending where the monitor is implemented, the service delivery paradigm can be different. In the case where the monitor is between the cloud OS and the hypervisor, the paradigm would be platform as a service. This is because the user can select the desired guest OS. However, there are multiple communication layers in the architectural stack. Should that model be implemented on a hardware level, the infrastructure as a service paradigm can be followed.

As the Single- and Multi-tenant VM models are very similar, their service delivery paradigm is also similar. In both cases the user can select the desired OS to deploy. Although, to the user it seems that the guest VM is deployed to the hypervisor, it is contained in the forensic VM. However, this does not change the service delivery paradigm. The platform as a service paradigm would thus be best suited for these models.

5. Summary

Table 1 summarizes the different models and compares them.

Table 1: Summary

	Operating System Embedded	Hypervisor Embedded	Communication layer	Single tenant VM	Multi-tenant VM
Implementation	Built into OS	Built into OS	Built into cloud OS / Application on cloud OS	Runs on Hypervisor	Runs on Hypervisor
Implementation Complexity	Medium	High	High	Low	Low
Data segregation	Per OS instance (High)	Managed by Hypervisor, dependent on implementation (best case: High)	Dependent on implementation. Part of cloud OS (Low - Medium). Part of hardware (Low)	Per OS instance (High)	Per Forensic OS instance (Medium)
Tamper resistance	Dependent on OS	High	Dependent on implementation. Embedded, High. Separate application, Low	High	High
As a Service model	SaaS	PaaS	IaaS/PaaS, depending on implementation	PaaS	PaaS

6. Future work

This paper only addressed the models in which a forensic monitor could be deployed in a cloud environment. Because of this limited scope, there are many points still to be addressed. These include the techniques that must be employed to capture relevant data via the monitor, how the captured data will be stored and how the data will be catalogued.

As the proposed models form part of a hosting solution, the business aspect cannot be ignored. Service level agreements would have to reflect what data is captured and for what purpose. Ownership of the data would have to be clearly defined. Linked closely to this, is the method which the captured data is stored and how the owner can be assured of both data integrity and privacy.

7. Conclusion

As vast cloud computing systems become more prevalent, the need to implement such models will become critical. It must be stressed that cloud service providers cannot run a viable business if the hosting systems have to be shut down for each investigation. The proposed models would negate this need as the information is already captured, catalogued and stored. The forensic readiness of cloud systems might then become another service in the offering of cloud service providers.

References

Andrew, M.W., (2007) Defining a Process Model for Forensic Analysis of Digital Devices and Storage Media. *Second International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'07)*, pp.16–30.

Ballou, S. et al., (2001) *Electronic Crime Scene Investigation: A Guide for First Responders*, Washington.

Chang, W.Y., Abu-Amara, H. & Sanford, J.F., (2010) Transforming Enterprise Cloud Services. In *Transforming Enterprise Cloud Services*. Springer, pp. 54–56.

Cohen, F., (2009) *Digital Forensic Evidence Examination* 3rd ed., Fred Cohen & Associates.

Hayes, B., (2008) Cloud computing. *Communications of the ACM*, 51(7), p.9.

Hess, K. & Newman, A., (2009) *Practical Virtualization Solutions: Virtualization from the Trenches*, Prentice Hall.

Kruse, W. & Heiser, J., (2002) *Computer Forensics: Incident Responce Essentials* 1st ed., Addison Wesley.

Lagar-Cavilla, H.A. et al., (2011) SnowFlock. *ACM Transactions on Computer Systems*, 29(1), pp.1–45.

Lillard, T., (2010) *Digital Forensics for Network, Internet, and Cloud Computing: A Forensic Evidence Guide for Moving Targets and Data* 1st ed., Syngress.

Manoel, E. et al., (2002) *Linux Clustering with CSM and GPFS*, IBM Redbooks.

Mell, P. & Grance, T., (2011) *The NIST Definition of Cloud Computing (Draft) Recommendations of the National Institute of Standards and Technology*,

Moreno-Vozmediano, R., Montero, R.S. & Llorente, I.M., (2009) Elastic management of cluster-based services in the cloud. In *Proceedings of the 1st workshop on Automated control for datacenters and clouds - ACDC '09*. New York, New York, USA: ACM Press, p. 19.

Nurmi, D. et al., (2009) The Eucalyptus Open-Source Cloud-Computing System. *2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, pp.124–131.

Pangalos, G., Ilioudis, C. & Pagkalos, I., (2010) The Importance of Corporate Forensic Readiness in the Information Security Framework. *2010 19th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises*, pp.12–16.

Perez-Botero, D., Szefer, J. & Lee, R.B., (2013) Characterizing hypervisor vulnerabilities in cloud computing servers. In *Proceedings of the 2013 international workshop on Security in cloud computing - Cloud Computing '13*. New York, New York, USA: ACM Press, p. 3.

Ras, D.J. & Olivier, M.S., (2012) Finding File Fragments in the Cloud. In G. Peterson & S. Shenoi, eds. *Advances in Digital Forensics VIII*. Springer Berlin / Heidelberg, pp. 169–185.

Rowlingson, R., (2004) A ten step process for forensic readiness. *International Journal of Digital Evidence*, 2(3), pp.1–28.

Ruan, K. et al., (2011) Cloud forensics. In *Advances in digital forensics VII*. Springer, pp. 35–46.

Sotomayor, B. et al., (2009) Virtual Infrastructure Management in Private and Hybrid Clouds. *IEEE Internet Computing*, 13(5), pp.14–22.

Van Staden, F.R. & Venter, H.S., (2011) Adding digital forensic readiness to electronic communication using a security monitoring tool. *2011 Information Security for South Africa*, (1), pp.1–5.

Szefer, J. & Lee, R.B., (2012) Architectural support for hypervisor-secure virtualization. *ACM SIGARCH Computer Architecture News*, 40(1), p.437.

Voorsluys, W., Broberg, J. & Buyya, R., (2011) Introduction to Cloud Computing. In *Introduction to Cloud Computing*. New York, USA, pp. 1–44.

Non Academic Paper

Distinguishing Cyber Espionage Activity to Prioritize Threats

John Hultquist
iSIGHT Partners, Chantilly, Virginia, USA
jhultquist@isightpartners.com

Abstract: The systems which run government and business are consistently under pressure from a multitude of adversaries with global origins, disparate capabilities, and a myriad of intentions and motivations. Adversaries differ tremendously in the resources they are willing to devote, their intentions for targeted systems, and their capability to affect these systems. Distinguishing between them is the key to effectively focusing limited security resources in protecting the enterprise. Distinguishing the activity of cyber espionage actors from other threats such as ecrime actors and hacktivists is complicated by the proliferation of tools as well as tactics, techniques, and procedures, the inherent difficulty of attribution, and the involvement of third parties in adversary operations, but discerning the adversary's motivation is a first step in drilling down to the fundamental questions of persistence, sophistication, and scope of threat. Despite the popular term "advanced persistent threat", much of cyber espionage activity is neither advanced nor persistent, and may challenge assumptions upon analysis. Many actors regularly leverage widely available tools and exploits to target low-hanging fruit. Furthermore, through the effects could be similar, in many cases strategically targeted operations are distinguishable from their tactical counterparts by their lack of focus and persistence against an organization. This paper discusses practical techniques for determining the motivation of the adversary based on historical cyber espionage activity.

Keywords: cyber intelligence, cyber espionage, APT, cyber threat intelligence

1. Introduction

Cyber espionage operators, characterized by their efforts to remain covert, can be distinguished from other threats by their unique characteristics and dissimilar behavior from hacktivists and cyber criminals. Recognizing these actors, who are often mischaracterized, is the key to making decisions with limited defensive resources and focusing on the offensive intent for targeted systems.

Any assessment of cyber espionage should first attempt to attribute actions to historic precedent, wherein actors have previously demonstrated their intent for systems, and possibly exposed their motivation. However, this scenario is ideal and defenders may not have access to such historic information. This paper attempts to outline characteristics which may aid in drawing conclusions without such long-term attribution.

2. Characterizing actors

2.1 Motivation

Cyberthreats principally emanate from three types of actors, characterized by their ultimate motivation for carrying out malicious activity:

- Cybercriminals
- Hacktivists
- Cyber Espionage Actors

Cybercriminals – Are principally engaged in or carrying out functions that enable fraud. These actors are motivated by financial recompense, in return for their part in the increasingly complex and specialized cybercrime marketplace. Typical cybercrime activity ranges from the development and deployment of tools for the widespread collection of credentials to the targeted intrusion of systems for similar data.

Hacktivists – These actors are principally ideologically motivated and carry out activity which is ultimately public in order to influence perceptions. In addition to ideological motivations, these actors may be motivated by ego, seeking to gain public recognition for themselves or their online personas.

Cyber Espionage Actors – Cyber espionage is the collection of information for the ultimate purpose of gaining information advantage, whether strategic or tactical, over the targeted entity. The activity, which is almost always surreptitious, may enable tactical insight through the collection of information such as negotiation positions or strategic advantage through the large scale collection of intellectual property. This activity is often popularly referred to as "APT".

Characterizing actors by motivation is an imperfect taxonomy, especially with regards to the ultimate sponsors of activity such as nation-states or non-state groups such as terrorists or corporations. However, this taxonomy recognizes the inherent complications of attribution, a hallmark of cyber threats.

2.2 Intent

In support of an ultimate motivation, which prompts action, actions can be characterized by their intent. Intent is a means to categorize actions by their relationship to motivation without overemphasizing the ephemeral tactics, techniques, and procedures which manifest them.

Cybercriminal Intent

- Compromise
- Exfiltrate
- Maintain Access
- Enable Fraud
- Disrupt

Hacktivist Intent

- Compromise
- Exfiltrate
- Destroy
- Disrupt
- Dump
- Deface
- Embarrass

Espionage Intent

- Compromise
- Exfiltrate
- Maintain Access
- Enable Information Advantage

3. The value of distinguishing actors

Enterprise security should consider the distinction between cyberthreats with the same objectivity that they consider the differences between their own organization and any other. Inherent disparities between an NGO and a multinational corporation, for example, may create distinctive threat profiles for each. With respect to the difference between actors, an organization might consider two factors:

3.1 Likelihood of being targeted by an actor with a given motivation

Actors pursue targets on the basis of their ability to contribute to their ultimate goal – influence, monetary gain, or information advantage. Several organizations offer a limited means to these ends. For instance, cyber espionage operators may be strongly incentivized to target dissident organizations, but that same organization may be a poor target for hacktivist operators asserting similar ideologies. Conversely, an oil and gas firm may find itself targeted by actors from both of these motivations; the former seeking intelligence on oil production while the latter carries out a disruption campaign for the ostensible purpose of environmental protection.

In addition to the value this factor has in preparing for threat scenarios, it can be considered to determine actor's probable long-interest following a compromise. Indiscriminate methods, such as watering hole compromises or spam may present malicious actors with the choice of whom to carry-out more resource intensive follow-on operations. Additionally, they may discover the target is at odds with their ideological

foundation and decline to carry out further malicious activity. This was the case in 2011 when LulzSec revealed a vulnerability to the United Kingdom's National Health Service, rather than exploit it (BBC 2011).

3.2 Intentional effects on an organization

Organizations have disparate tolerances for brand damage and monetary loss, as well as threats to secrecy, legitimacy, customer confidence and the loss of competitive advantage. A government contractor involved in research and development may find disruptive threats to their public-facing website from hacktivists inconsequential, but loss of their intellectual property existential. Alternatively, a retail organization may suffer a loss of customer confidence from any public event, whether its hacktivist or criminal in nature.

In some cases, the effects of cyberthreats can be measured relative to antipode investments. For instance, the antipode to a loss of intellectual property is the outcome of research and development. Marketing investment is the antipode of brand damage that might result from an incident.

4. Actions on objective

Obviously, intent is most clearly indicated by the actions adversaries take or have-taken previously, if attribution is possible. However, a long term objective may be obscured by intermediate steps. For instance, the actions of an adversary seeking to maintain long-term access to systems for the ultimate goal of gathering intelligence may look entirely similar to the actions of an actor who intends to publicly expose compromised information. Both actors will seek to gain access and maintain or expand it until they compromise the information they seek.

Furthermore, it may be impossible to gain insight into final steps taken by the actor. Cybercriminal and hacktivist actors may expose their intent when compromised data is used or dumped, but the nature of espionage is far more surreptitious, and the beneficiaries of it are more likely to conceal sources and methods.

Prior to final steps, it may be difficult to distinguish the actor's ultimate intent, however, other factors, such as persistence on the objective over time and the type of information collected, may be indicative.

4.1 Persistence over time

Cyber espionage actors typically maintain access in perpetuity in order to continue to collect intelligence on the objective. As long as the target offers the actor the possibility of new information they will remain incentivized to maintain access. This is especially the case with strategic objectives, such as decision makers or intellectual property creators, which offer long-term benefits to the collector. Cyber espionage intrusions have lasted as long as five years.

By contrast, the public nature of hacktivist actions is a competing priority for long-term persistence. A public dump or leak will usually expose the source, ending the hacktivists access. However, it should be noted that actors seeking to expose data may lack the means to do so causing actors to maintain access for the long term.

Commodity cyber-crime activity has a marketplace incentive to turn-over control of compromised systems following successful compromise. In the highly-specialized underground, actors are willing to pay for systems which have already been subject to credential compromise for alternative purposes, such as spam or DDoS capability. Though general threats may persist under these conditions, the original actor may have relinquished control and the malware used by the author replaced or supplemented with other tools.

4.2 Type of information collected

The nature of information targeted by an intrusion is indicative of the actor's goal. Espionage operators typically seek access to documents, email, contacts, and communications systems which may offer similar opportunities and several implants employed by these actors are purpose-built to compromise this information. By contrast, cybercriminals, especially those employing commodity malware, will use a variety of tools purpose-built to collect monetizable information such as credentials, financial and personally identifiable information.

Information collected by hacktivists may resemble either cybercriminal or cyber espionage activity. Given that hacktivists compromise information for public release, rather than the intrinsic utility of the data they select information that will influence an audience. Previously this has included both financial information, credentials, and sensitive internal information such as intellectual property. AntiSec actors released credit card information, passwords, and internal correspondence as part of their raids against Stratfor (Norton 2011).

5. Targeting

Cyber espionage activity is by nature targeted, though it is not alone in this regard. Cybercrime and hacktivist activity may be targeted as well. Though cybercriminals employ commodity malware in large-scale indiscriminate activity, alternative approaches are often necessary for attempts to collect specific, discreet datasets, such as bulk financial information compromised from a retailer. Hacktivists have carried out several targeted actions, including intrusions, especially in response to ideological conflicts with targeted organizations.

Determining whether activity is targeted involves assessing information available to the adversary, and the adversary's probable thought-processes from their choices in conducting the intrusion attempt. The following factors should be considered:

5.1 Function or role of targeted personnel

The target of an incident represents specific value to the cyber espionage adversary and is selected by its perceived access to sensitive information or connection to other personnel or systems with such access. Personnel whom the adversary perceives represent these opportunities are simply more likely to be targeted than others. Someone involved in sensitive work or involved in high-stake decision making, for instance, represents a greater opportunity than others and their targeting may be indicative of a specific interest. However, the operator may have limited information available to them and must work within these confines. Therefore, the value of targeted personnel should be considered within the confines of information available to the adversary.

5.2 Source and exclusivity of target information

If targeted entities have limited the exposure of their contact information, any source for that information may illuminate the adversary's motivation. For instance, if the contact information is limited to a single source, such as the attendee list of an aerospace conference, this may expose an adversary interest in aerospace. A greater cohort of targets taken from such a source further emphasizes adversary use of the information, underscoring possible adversary interest.

5.3 Social engineering

Attempts to manipulate the target convey the probable scope of an operation as well as the adversary's intended target population. Social engineering with wide appeal such as emergency notices are common in cybercrime operations targeting large cohorts. By comparison, cyber espionage operations typically employ messages selected for a specific audience. A long-term espionage campaign against the US Government spoofed internal communications from the highest authorities to entice victims to open malicious documents containing the Zeus Trojan (Krebs 2010).

5.4 Watering hole intrusions

Strategic web compromises, also known as watering hole intrusions, are carried out when cyber espionage actors leverage a specially selected and compromised website to deliver exploits to the websites regular users. Sites chosen for this method expose motivation. For instance, several sites popular with dissidents such as the website of Amnesty International have been compromised in a watering hole campaigns almost certainly intended to monitor dissident activity (Bennett et al 2013).

6. Malware capability

Malware employed by the adversary is often used as a sole means of determining attribution to cyber espionage actors. In fact, the names of many commercially available malware signatures contain the designation "APT", intimating a relationship between the code for which the signatures are based and the

motivation behind the code's user. While malware is a useful tool in determining attribution, several additional factors should be weighed:

6.1 Historical use of the malware

Previous use of malware is a limited guide in distinguishing the activity of cyber espionage actors. Several cyber espionage actors use proprietary malware and though these actors have access to the cybercrime marketplace, they often eschew commodity malware for custom implants. However, there are several widely used variants employed by espionage actors, such as Poison Ivy RAT, Gh0st RAT, and Xtreme RAT. Any consideration of previous malware use should consider access to that malware.

6.2 Propagation and access to the malware

Poison Ivy RAT is available through a website which includes support for the relatively simple tool. The extremely reduced barrier to entry for the tool, which has been likened to an AK-47, invalidates any assumptions regarding motivation based on its use alone. By contrast, several tools are developed in closed environments and not seen outside of espionage or cybercrime. PlugX, a Chinese tool regularly used in espionage, is not widely available and to date has not been attributed to any criminal activity. Given these factors, the use of PlugX is a strong indicator of cyber espionage, though not a perfect one, due to limits in visibility.

6.3 Malware functionality

If propagation of malware is limited, its functionality may indicate an actor's intent. Whereas cybercrime malware is often purpose-built to intercept and collect financial data, cyber espionage malware focuses on principally on documents. Nettraveler malware, for instance, exfiltrates Adobe Acrobat and Microsoft Office file types as well as AutoCAD project files. The later file type would have little value to an actor seeking to monetize financial information (Kaspersky 2013.

7. Exploits

Similar to tools, exploit code is only useful in attribution to a given motivation so long as its propagation is limited. Following mainstream discovery of a zero-day exploit, the capability is typically adopted within days by exploit pack developers and open exploit frameworks such as Metasploit, diminishing any distinctive quality to its use. However, prior to the zero-day event, the exploit code can be considered novel and distinctive to a limited set of actors with access to the exploit's developers. This phenomenon owes to the resources necessary to develop or purchase exploit code and its value as a surreptitious capability. Following exposure, the value to those using the exploit as well as those who sell and develop them is entirely diminished, incentivizing these actors to limit its propagation.

Access to zero-day exploits is often regarded as a measure of sophistication. Because there is an active marketplace, and they can be sold, they are more accurately described as a measure of resources. Their development takes considerable time and capability, which can be roughly valued in monetary terms. Zero-day exploits can be sold for tens of thousands of dollars in the cyber crime underground (Greenberg 2012.

Access to zero-day exploits is not in itself indicative of intent, though their use in non-covert activity would be an injudicious use of this capability. For instance, a data dump carried out by hacktivist actor, would risk exposing an exploit, if it were used in obtaining the data. Several hacktivist actors, such as AntiSec and Syrian Electronic Army have had success without relying on zero-day exploits, preferring instead to rely on known exploits and social engineering techniques.

Furthermore, because of their value, the use of zero-day exploits may be injudicious. The use of a zero-day risks its exposure, incentivizing actors to use alternatives when they are sufficient. Cyber espionage actors regularly employ well-known exploits such as CVE-2012-0I58, despite long term awareness of the issue (Bennett et al 2013).

8. Infrastructure

Infrastructure used by cyber espionage operators is characterized by its relatively limited use in comparison to alternatives employed in commodity cyber crime. The latter is often designed with discovery in mind, and

developed to remain viable in spite of mitigation attempts. To this end, infrastructure is associated with domain generation algorithms, fast flux, and bullet-proof hosting, schemes which prolong the activity in spite of intervention. Infrastructure of this nature is often vast and complex.

Cyber espionage operators regularly label subdomains after targeted organizations in an attempt to conceal traffic from victim systems. This method creates a record of targeting in infrastructure, indicating the expanse of adversary interest. For example, the use of consillium and voanews as subdomains indicate targeting of European Union and Voice of America, consistent with geopolitical motivation.

9. Conclusion

Because few indicators explicitly identify an actor's motivation, each should be judged by the probability that it indicates cyber espionage. Conversely, the same indicators may suggest alternatives – hacktivism or cyber crime, and any ultimate judgment should give these alternatives equal consideration. Final judgments should consider the probability that each factor indicates a given motivation, and weigh these factors in total to determine the most likely scenario.

Changes in technology, behavior, or other unforeseen circumstances are almost certain to cause actors to eschew current norms. Careful analysis must always consider the previously unseen, especially considering the secretive nature of the cyber espionage actor.

References

Bennett, James T., Ned Moran, and Nart Villeneuve *Poison Ivy: Assessing Damage and Extracting Intelligence* 2013 Fireeye http://www.fireeye.com/resources/pdfs/fireeye-poison-ivy-report.pdf

British Broadcasting Corporation *Hackers warn NHS over security* BBC.com 9 June 2011 http://www.bbc.co.uk/news/technology-13712377

Greenberg, Andy *Shopping For Zero-Days: A Price List For Hackers' Secret Software Exploits* 23 March 2012 Forbes.com http://www.forbes.com/sites/andygreenberg/2012/03/23/shopping-for-zero-days-an-price-list-for-hackers-secret-software-exploits/

Kaspersky Global Research and Analysis Team *The NeTTraveler (aka 'TravNeT')* June 4, 2013 http://www.securelist.com/en/downloads/vlpdfs/kaspersky-the-net-traveler-part1-final.pdf

Krebs, Brian *Zeus Attack Spoofs NSA, Targets .gov and .mil* Krebs on Security 6 February 10 http://krebsonsecurity.com/2010/02/zeus-attack-spoofs-nsa-targets-gov-and-mil/

Norton, Quinn *Antisec Hits Private Intel Firm; Millions of Docs Allegedly Lifted* Wired.com 26 December 2011 http://www.wired.com/threatlevel/2011/12/antisec-hits-private-intel-firm-million-of-docs-allegedly-lifted/

Work in Progress Paper

NATO Article Statue 5 in Terms of a Cyber-War

Selcuk Dal and Kadir Ozyurt
Turkish Army War College, Istanbul, Turkey
selcuk2001@gmail.com
kadirozyurt97@gmail.com

Abstract: Nowadays, cyber-attacks occur every day in numerous numbers. Cyber threat is real and countries struggle to do their best to be prepared to cope with them. Most of the NATO members are also victims of those attacks. Accordingly, NATO is trying to increase its readiness to win this new type of war. Fortunately, except for Estonia (Mansfield, 2012), there has not been a large scale cyber-attack to any NATO member country, yet. However, this doesn't necessarily mean that it is unlikely. Secondly, NATO article 5 has never been invoked as a cyber response option in the Alliance's history. In case of a large scale cyber-attack to a NATO country or countries, can the Article 5 be invoked? If yes, to what degree? May active cyber defense be a method to implement Article 5? This paper tries to find out whether article 5 can be implemented in case of a cyber-attack to a member country and what can be done in retaliatory actions (Tape, 2010). Should the response to a cyber-attack be a counter-cyber-attack or could a full-scale use of force be an option?

Keywords: NATO, Washington Treaty Article 5, cyber warfare, threshold of a cyber attack

1. Introduction

NATO is an organization that safeguards the freedom and security of its members by political and military means (A political and military alliance, n.d). Although it is a peaceful organization, when diplomatic efforts fail, NATO has the military capacity to undertake crisis-management operations. One of the important articles of Washington Treaty – founding treaty of NATO - is the Article 5. According to Article 5 (North Atlantic Treaty 1949):

> *The Parties agree that an armed attack against one or more of them in Europe or North America shall be considered an attack against them all and consequently they agree that, if such an armed attack occurs, each of them, in exercise of the right of individual or collective self-defence recognised by Article 51 of the Charter of the United Nations, will assist the Party or Parties so attacked by taking forthwith, individually and in concert with the other Parties, such action as it deems necessary, including the use of armed force, to restore and maintain the security of the North Atlantic area. Any such armed attack and all measures taken as a result thereof shall immediately be reported to the Security Council. Such measures shall be terminated when the Security Council has taken the measures necessary to restore and maintain international peace and security.*

Article 5 is one of the most important features of the Washington Treaty since it ensures that all members agree to help and use armed force when necessary. In the cold war era this meant a lot and kept the alliance together. However, in this age, threats are much more different than that of 1950's. For example, in 2006 Riga Summit, cyber-attacks were deemed among asymmetric threat (NATO, 2006). To encounter those threats in accordance with Article 5 is a question that remains in this era.

Cyber threats are among the most important threats of this century. As long as cyber-attacks have prevailed and threatened the countries' security, NATO members developed their cyber defense capabilities. A lot of efforts have been put forward to improve those capabilities. Today, NATO is in a better position in terms of cyber capabilities compared to that of 1990's. After the Estonia cyber-attack, NATO made some noteworthy progress as to its cyber defense capabilities. First, The Cooperative Cyber Defense Center of Excellence (CCD CoE) was created in 2008 to enhance the capability, cooperation and information sharing among NATO, NATO nations and partners in cyber defense by virtue of education, research and development, lessons learned and consultation (Myrli, 2009). The main tasks of CCD CoE are (Bart, 2013); providing cyber-related doctrines and concepts for the Alliance, hosting and conducting training workshops, courses, and exercises for NATO member states, conducting research and development activities, studying past or ongoing attacks to draw up lessons learned, and providing advice, if asked, during ongoing attacks.

Although NATO members are exposed to cyber-attacks on a daily basis, it has never been considered if Article 5 can be invoked against the attackers. However, when the rising number of cyber-attacks is considered, invoking Article 5 remains as an important question that should be examined thoroughly. While answering the

question of whether Article 5 can be invoked in case of a cyber-attack, there are some important things to point out (Figure 1): the attribution problem, cyberspace definition and threshold of an armed attack. In addition, since cyber warfare is a new warfare method, Article 36 of 1977 Additional Protocol I to the 1949 Geneva Conventions should also be considered.

Figure 1: Problems of invoking Article 5 in cyber attacks

2. Attribution problem

First of all there is an attribution problem in cyber-attacks. It is not easy to say who is behind a cyber-attack since it is relatively easy to hide the evidences. The most important piece of evidence is to find the IP number of attacker. However, you may not be able to find an IP number after an attack, or the IP number you find may be misleading. IP number can be false or may be rerouted to various locations. Another thing to look at in case of a cyber-attack is the structure of malware. The language used in the malware (Arabic, Russian, Chinese etc.) and the time it was compiled tell a lot about the attacker. Yet, this information can easily be forged, as well. Thus, attribution remains difficult in dealing with the attacker (Shakarian et al, 2013).

3. Cyberspace definition

Secondly, cyber-attacks are not always conducted by states. Criminals and even some computer geeks may be behind cyber-attacks. Of course it is not reasonable and possible to invoke Article 5 against some criminals or geeks. Mälksoo (2013) states that even if there is a state behind a cyber-attack there are two requisites to invoke state responsibility:

- cyber-attackers could be qualified as actors supported by state A (or as de facto representatives of the state, in case the respective support is conclusively proven);
- or, it is successfully proven that state A government was standing right behind the cyber-attacks organized against state B.

In addition, since there is no distinction between national, international, public or private layers in cyberspace, the cyberspace that should be defended is disputable. Moreover, it is impossible to defend even any specific area in cyberspace due to the plethora of things to defend but only a single vulnerability is enough to be exploited (Mälksoo: 2013). Hence, due to the complex nature of cyberspace and cyber threats, it is difficult to invoke Article 5.

4. Threshold of an armed attack

Another issue is determining the threshold that will trigger Article 5. Since Article 5 states that there should be an “armed attack” for the NATO member to have a self-defense right according to Article 51 of the Charter of United Nations, a physical mischief is a necessity in this case. As the existing international law on the use of force does not actually prescribe which means could be used by a state for its self-defense – as long as these means are proportional to the extent of the damage caused by an attack on a state and inevitable in order to retaliate the attack, it would follow that a victim of a de facto attack could legitimately use military force to retaliate cyber-attacks (or any other attacks of the non-traditional kind) that have caused its structures and people real and considerable harm, as well (Mälksoo: 2013). Scmitt (2010) states that a cyber-attack standing

alone will comprise an armed attack when the consequence threshold is reached. Equally, states subjected to an armed attack may elect to respond solely with cyber operations. However, it seems, it is not possible to set a general rule for the cyber-attacks since the effects of the cyber-attacks should be evaluated case by case.

Given all the issues above, invoking Article 5 seems quite problematic indeed. Yet, under any circumstances where a cyber-attack incurs concrete harms as in physical war, attribution is successfully made, and the attacker is a state or a state is responsible for the attacks, Article 5 seems robust enough for a partial and/or full campaign. In this case all response options including military ones can be employed against the attacker. Among these options cyber-attack may take place as well.

5. Article 36

Article 36 of 1977 Additional Protocol I to the 1949 Geneva Conventions states that:

> *In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.*

Obviously, cyber warfare is a fairly new warfare. Hence, issues about new warfare falls well under this article. Thus by looking at Article 36 it can be concluded that applying cyber warfare should not be against any of the international law terms. For example, Article 56 states:

> *Works or installations containing dangerous forces, namely dams, dykes and nuclear electrical generating stations, shall not be made the object of attack, even where these objects are military objectives, if such attack may cause the release of dangerous forces and consequent severe losses among the civilian population. Other military objectives located at or in the vicinity of these works or installations shall not be made the object of attack if such attack may cause the release of dangerous forces from the works or installations and consequent severe losses among the civilian population.*

However, experts remind the possibility of a cyber war, with enemy states exploding fuel refineries or sabotaging air traffic control systems (Blitz:2013). Unfortunately, in brief history of cyber attacks, a number of cyber attacks against dams and nuclear power stations has appeared. As Article 56 states, it is prohibited to attack such facilities. Cyber attacks is not an exception. But, since attribution is highly difficult in cyber attacks, such facilities are exposed to cyber attacks. Yet, one should know that this is against the international law.

6. Conclusion

NATO has always been an organization that guarantees the security of its members by political and military means. It has successfully accomplished its core tasks by employing relevant response options. Yet, in this era in which threats are very different from that of the previous age, the organization encounters several impediments in replying the new kinds of threats. One type of those threats, cyber threat remains as an important thorny issue to be clarified.

NATO has accomplished quite well set of things to be better off in this new cyber threated environment. The Cooperative Cyber Defense Center of Excellence is a good example of those efforts. Every day it is trying to be more prepared to a cyber-war. However, a question still remains to be answered: in case of a cyber-war, whether article 5 can be and to what degree invoked?

In the nature of cyber war there lies *some* difficulties. For example, how to make the attribution, how to define and defend the cyberspace, and determine the threshold of an armed attack. In addition, since cyber warfare is a relatively new warfare, cyber attack planners should follow article 36 of I977 Additional Protocol I to the 1949 Geneva Conventions and nothing should conradict international law. In conclusion, Washington Treaty Article 5 seems to be invokable under the conditions that a successful attribution is made, it is proved that the enemy is a state or a state is responsible of the attacks, and a physical mischief occurs.

References

A political and military alliance n.d. Available from: < http://www.nato.int/nato-welcome/index.html>. [25 December 2013].

Blitz,JAmes, 2013, "Global politics: Fears of cyber war and espionage raise tensions", Financial *Times 23* April. Available from <http://www.ft.com/cms/s/0/e758e548-a3e6-11e2-ac00-00144feabdc0.html#axzz2ySAnceBr>. [23 April 2013].

Mälksoo, Mario (2010) "Article 5 and the Strategic Concept", Proceedings of The future of NATO's Nuclear Deterrent: The New Strategic Concept and the 2010 NPT Review Conference. Available from: < http://www.icds.ee/fileadmin/failid/Maria_Malksoo_on_cyber_and_energy_security_and_Article_5.pdf> [18 January 2014].

Mansfield, Steve, (2012) "Estonia:What doesn't kill you make you stronger", Network security, 2012(7), pp.12-20.

Myrli, Sverre, NATO Parliamentary Assembly, "173 DSCFC 09 E bis - NATO and Cyber Defence" Available from: <http://www.nato-pa.int/default.asp?SHORTCUT=1782>. [18 August 2013]. NATO Press Release, Riga Summit Declaration (2006), Available from: http://www.nato.int/docu/pr/2006/p06-150e.htm. [18 January 2013].

North Atlantic Treaty, Washington D.C. (1949), Available from: <http://www.nato.int/cps/en/natolive/official_texts_17120.htm > [25 December 2013].

Schmitt, Michael N. (2010) "Cyber Operations in International Law: The Use of Force, Collective Security, Self-Defense, and Armed Conflicts" Proceedings of a Workshop on Deterring CyberAttacks: Informing Strategies and Developing Options for U.S. Policy, pp.151-178. Available from: < http://www.nap.edu/openbook.php?record_id=12997&page=151> [21 January 2014].

Shakarian, Paulo, Shakarian, Jana & Ruef, Andrew (2013) "Introduction to cyber-warfare", Elsevier, Waltham.

Smedts, Bart (2010) "NATO's Critical Infrastructure Protection and Cyber Defence." *Royal High Institute for Defence Center for Security and Defence Studies*, Brussels.

Tape, Red (2010) "Could cyber skirmish lead U.S to war?", Available from: <http://www.nbcnews.com/technology/could-cyber-skirmish-lead-u-s-war-6C10406234>. [21 January 2014].

Late Submission

Human Factor: The Weakest Link of Security?

Roman Jasek and Libor Sarga
Tomas Bata University in Zlín, Zlín, Czech Republic
jasek@fai.utb.cz
sarga@fame.utb.cz

Abstract: Human element plays a critical role in cyberwarfare scenarios: a malicious adversary can launch targeted social engineering campaigns to gain unfettered access to sensitive electronic resources, establish unauthorized system persistence, and use the compromised host as a stepping stone for further exploitation, incorporating it into a botnet of controlled nodes. As hardware and software infrastructure protection efforts result in increasingly resilient systems, focus on end-users who constitute a security vulnerability can be expected to increase in the future. However, password database leaks, effectiveness of social engineering, and bring your own device (BYOD) trends in organizations all raise concerns as to the security competencies the general population possess. In the article, we present results of a large-scale questionnaire study pertaining to security habits and BYOD practices of more than 700 participants conducted in the Czech Republic during the period of September—December 2013. Ranging from a preferred operating system to password selection rationale, the answers should be a representative cross-section of how an "average" user maintains their electronic identity online. The snapshot provides valuable insights and actionable intelligence based on which information and communication technology policies in organizations can be modified to better accommodate the patterns discovered. The article maps current state of selected aspects of security in increasingly interconnected, technology-driven global structures where electronic identities supplement real-world ones and their compromise results in significant negative consequences.

Keywords: research, study, questionnaire, BYOD, password, security

1. Introduction

Information and communication technology (ICT) is a cornerstone of modern economies, financial markets, entertainment, medicine, politics, and scientific applications, enabling large-scale data processing and real-time results presentation. Networks allow people and organizations to connect across geographical boundaries, obviating the need for physical presence while fostering productivity and cost effectiveness. A majority of corporations and public administrations adopted ICT to process various types of sensitive records and communicate with clients and citizens.

However, due to its increasing complexity, ICT is nowadays an extensive system of interrelations between hardware, software, people, and processes each of which suffers from weaknesses third parties can exploit and circumvent their intended functionality. Recognizing this, countermeasures have been put in place designed to mitigate or eliminate exploitable vulnerabilities. While we argue hardware, software, and processes can be modified to a high degree, the human factor is notoriously error-prone, its mindset almost impervious to change in short term, and behavior patterns by which it is governed unpredictable. Furthermore, as programmatic defensive measures have become more sophisticated, people are increasingly perceived as the single weakest link of security. Systematic research in this area is therefore warranted in order to understand motives and real-world practices of users, particularly in face of emerging trends such as Bring Your Own Device (BYOD) which sees integration of user-owned mobile devices with corporate ICT infrastructures.

In the article, we present a large-scale questionnaire research conducted during the period of September—December 2013 which polled users on various aspects of ICT. Employing offline data collection methodology, the results are not limited to Internet-proficient respondents, but are based on a cross-section of general population instead. It is structured as follows: section 2 provides brief overview of BYOD and password management, two factors which will be preferentially scrutinized. Sections 3 contains main contributions including findings and viable attack vectors. Section 4 offers concluding remarks.

We believe the research contributes to user-side security with actionable intelligence for organizations attempting to improve their ICT resilience by addressing vulnerabilities in the human element.

2. 2. Selected Aspects of ICT Security

In this section, we briefly examine selected aspects of two trends pertinent to ICT security. The first one emerged due to rapid adoption of smart mobile devices and their interactions with sensitive data in organizations, the second one is an ongoing issue where a balance needs to be established between user comfort and security.

2.1 BYOD

The idea of ubiquitous computing as unrestricted, pervasive electronic resource availability, networks of interconnected devices, and scalability was envisioned as "…[t]he idea of integrating computers seamlessly into the world at large…" (Weiser 1991). A more recent delimitation sees it as utilizing "…countless very small, wirelessly intercommunicating microprocessors, which can be more or less invisibly embedded into objects" (Friedewald and Raabe, 2011). One device class in particular has brought ubiquitous computing to consumers: small form factor devices, specifically smartphones and tablets. It was noted that "…[t]he emerging capabilities of smart phones are fueling the rise in the use of mobile phones as input devices to the resources available in the environment… The ubiquity of mobile phones gives them great potential to be the default physical interface for ubiquitous computing applications" (Ballagas et al, 2006). We will further discuss smartphones because their ubiquity among users presents a challenge for effective security management.

Growth of smartphone ecosystem led to the introduction of BYOD (Bring Your Own Device) or more recently, BYOT (Bring Your Own Technology), subsets of consumerization aimed at mobile hardware (Scarfo, 2012) and hardware with software, respectively. Apart from privacy issues, security challenges were pointed out as well, and "[w]ith the advent of cloud storage with its partitions, care should be taken in-house to ensure that data is partitioned and individual users only get access to the information they need to perform their assigned duties" (Miller, Voas, and Hurlburt 2012). The authors also admit that "…little attention has been paid to this issue, but that's a problem that will need to be addressed if BYOD and BYOT become adopted widely…" Irreversible modifications are not an option which "…stems from the fact that given the device does not belong to the enterprise, the latter does not have any justification – and rightly so – in modifying the underlying kernel of the personal device of the employee" (Gessner et al, 2013).

Mobile cyberwarfare has thus been in the focus of the security community and is believed to be the next stage where the asymmetric conflict between attackers and defenders, the former attempting to penetrate the system by finding an exploitable vector, the latter attempting to cover every possible venue, will take place (Sarga, Jašek, 2013). User participation is necessary as smartphone functionality becomes on par with desktop stations.

One way to manage disparate mobile devices is through profiles. We understand profiles as collections of permissions and restrictions installed on the device which can be utilized for setting mandatory policies when remotely accessing information systems processing sensitive data. Companies use them in situations where permanent modifications cannot be utilized because the devices belong to employees. By deploying a means which ensures best security practices (encryption, password composition) are met when accessing internal electronic assets, the perpetrator cannot employ techniques such as man-in-the-middle (MITM) where data are passively intercepted or actively modified when passing through a malicious node, the result of connecting to unsecured wireless networks.

2.2 Password Management

Passwords are common, intuitive means of authentication. However, "[a]lthough the user selects a password by combining characters or numbers than can be selected from the keyboard, [passwords consisting] of consecutive numbers, specific words or sentences are frequently used for the most part" (Kim, 2012). The practice is not exclusive to mobile devices where convenience and typing speed is preferred over complexity but also in situations where full keyboard is available. Empirical findings on patterns found in passwords (Klein, 1990; Zviran and Haga, 1999; Yampolskiy, 2006) have corroborated the hypothesis "…people's choice of passwords is non-uniform, leading to some passwords appearing with a high frequency…. [O]ne consequence of this: a relatively small number of words can have a high probability of matching a user's password. To combat this, sites often ban dictionary words or common passwords… in an effort to drive users away from more common passwords" (Malone and Maher, 2012).

As password choices were found strongly non-uniform and word lists can be used to compile likely candidates, brute-force enumeration covering the whole search space and assuming each string (password) is equally likely to occur has been refined and optimized into a dictionary attack. Defined as "[a]n attack that uses a brute-force technique of successively trying all the words in some large, exhaustive list," (Shirey, 2007) it increases search space without resorting to large-scale enumeration by employing a sophisticated set of rules to mutate the strings in the source dictionary.

Organizations can mandate minimum password composition requirements to ensure the work factor for any malicious third party is prohibitively high, supposing a strong one-way cryptographic hash function has been used to convert the string. The policies should vary for desktops with dedicated input peripherals and smartphones employing touch interfaces and virtual keyboards on screen. Yan et al (2004) state the following: "Good password appear to be random characters. The wider the variety of characters the better. Mixing letters with numbers is better than letters alone. Mixing special characters with number and letters is better still. One recommendation that seems increasingly popular is the pass phrase approach to password generation." Empirical results favor passwords of length at least 16 (Kelley, 2012) consisting of non-repeating characters which do not form discernible patterns, or can be found in a dictionary. Password length should increase with advances in hardware performance, parallel computations, and algorithm design.

We argue decoupling users from security to the highest extent possible is warranted. The concept, titled security as a service (SECaaS), is based on the assumption low-level details should be hidden from users who request a service, e.g., strong password, and the system provides it with minimum delay. Generating login credentials coupled with dedicated software storing them in encrypted containers is an example of how organizations can assure high level of protection while maintaining user comfort. Each employee is asked to remember a single string (alternatively, biometric authentication can be used) to access their database from which passwords can be copied, precluding the need for memorization.

Three factors influencing password security are complexity, length, and uniqueness. Brute-force attacks exploit short length by covering the search space exhaustively. At the expense of long execution time, it is guaranteed to find the password every time; for longer strings, dedicated hardware components with larger throughputs are preferred over central processing units (CPU). Hardware being currently capable to enumerate all passwords up to and including 8 characters, the bound will increase which should be reflected in the composition policies. Dictionary attacks exploit low uniqueness where word list entries are copied or slightly modified, e.g., substituting characters, prepending/appending symbols, or concatenating several words. The schemes do little to thwart even low-skilled adversaries who can employ simple rules designed to reveal the original string.

3. Questionnaire Research Findings

Data for the research was sourced from Master degree students at the Faculty of Management and Economics, Tomas Bata University in Zlin in both full-time and distance forms, each was instructed to distribute and return 10 copies in paper form. Questionnaires were handed over in September 2013 and returned during the period of October—December 2013. The sample exclusively comprises citizens of the Czech Republic. The data were manually typed into a Microsoft Excel spreadsheet, imported and processed using IBM SPSS Statistics 22 which generated all tables and graphs presented in the article.

An important future research venue is to extend the research to other countries, and thus validate the results which are currently applicable only to the Czech Republic on a larger statistical sample. Even though we hypothesize no significant differences will be found, mapping user sentiments across geographic boundaries ensures the findings are applicable globally instead of locally.

The questionnaire consisted of four parts: general IT overview, mobile phones, additional questions, and personal information. The two categories presented here are general IT overview and mobile phones. We first present gender and age structure of respondents, denoted as q4.1 and q4.2, respectively. Figure 1 demonstrates the compositions.

Statistics

		q4.1	q4.2
N	Valid	784	783
	Missing	0	1

q4.1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	422	53,8	53,8	53,8
	Male	362	46,2	46,2	100,0
	Total	784	100,0	100,0	

q4.2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<18	55	7,0	7,0	7,0
	19-25	318	40,6	40,6	47,6
	26-35	170	21,7	21,7	69,3
	36-45	78	9,9	10,0	79,3
	46-55	118	15,1	15,1	94,4
	56-65	22	2,8	2,8	97,2
	65>	22	2,8	2,8	100,0
	Total	783	99,9	100,0	
Missing	999	1	,1		
Total		784	100,0		

Figure 1: Age and gender composition of the representative sample (Own work)

Three age groups were more frequently included compared to others, namely 19—25 which suggests students filled out the questionnaire themselves, 26—35, and 46—55. The results are therefore mainly indicative of the opinions and practices of the three most populous groups, providing insight into how they perceive and apply security principles in real-world situations.

3.1 BYOD

Out of 782 valid answers, 518 respondents (66.2 %) acknowledged having a smartphone or a tablet, 264 (33.8 %) answered negatively. Though the research was a one-time event, conducting the survey at multiple reference points would allow for analyzing whether the adoption rates increase, stagnate, or decrease over time. Assuming the general population gradually migrate towards newer technologies, it can be hypothesized smartphone share will rise, especially as cheaper models are promoted which justify the cost for consumers whose needs have so far been satisfied by feature phones. Mobile operating system preference is depicted in Figure 2. The breakdown is as follows: from 730 valid answers, 221 (30.3 %) respondents indicated no preference, 341 favored Android (46.7 %), 98 iOS (13.4 %), 65 Windows Phone (8.9 %), and 5 other operating systems (0.7 %), specifically BlackBerry.

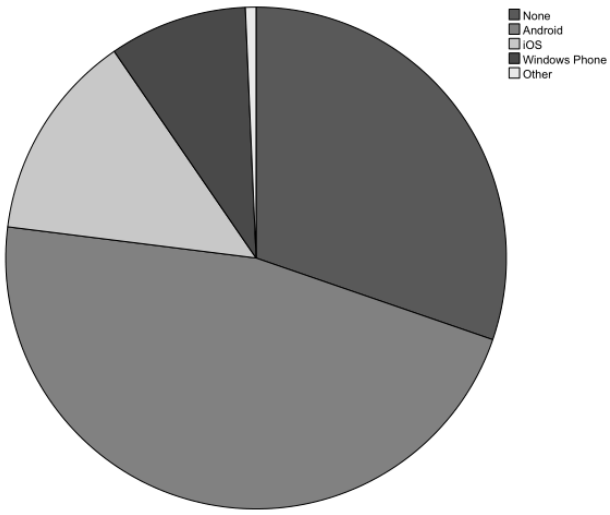

Figure 2: Mobile operating systems preferences (Own work)

Android is the most popular mobile operating system with almost half of research participants preferring it. We claim broad portfolio of low-cost, affordable devices contribute to the results. Compared with iOS found on relatively high-priced devices, the open-source nature of Android allow customizations and deployment on heterogeneous hardware setups. This fosters user adoption but also leads to version erosion and security risks because vendors routinely discontinue support for their products, leaving them open to vulnerabilities discovered and patched in later iterations. The fact was particularly noticeable on Android 2.3.x which lacked many security features mitigating execution of malware.

As for why the particular mobile operating system, and in extension the platform, is favored, respondents were asked for specific reasons. From 643 valid answers, the following were listed as the most important:

- applications (117, 18.3 %),
- it was a gift (110, 17.2 %),
- aesthetics (52, 8.1 %),
- previous experience (50, 7.8 %),
- brand (44, 6.9 %),
- recommendations from relatives or friends (44, 6.9 %),
- combination of aesthetics, applications, and brand (23, 3.6 %),
- switch from another platform (22, 3.4 %),
- other reason (18, 2.8 %).

Application ecosystem proves to be the deciding factor when choosing a phone. The second most-cited reason, gift, suggest respondents do not seek particular brand but are content with whatever smartphone they are given as their fundamental capabilities are largely unchanged. We believe consumers switching from feature phones would belong in this category, or alternatively would be influenced by recommendations from relatives or friends, preferring identical user experience. Aesthetics and previous experience were represented almost identically, brand was also cited as somewhat relevant, hinting at certain socio-economic status or non-explicit associations tied with owing a particular brand. Design philosophy apparently sees some users inclining towards tightly-controlled touch interfaces while others value broad customization. Switch from another platform could either suggest dissatisfaction or willingness for experimentation. Among other reasons, open-source nature and price/quality ratio dominated; price was not singled out, strongly suggesting it is secondary for the research participants.

The scope of smartphone use is demonstrated in Figure 3. Each main option was assigned a numeric code (1—5) and combinations were denoted as 1+2, 1+3, etc.

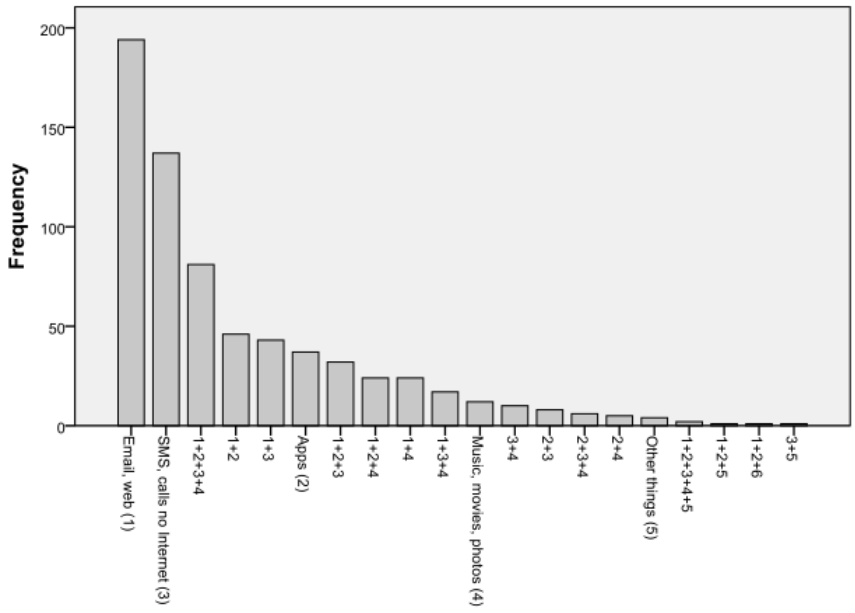

Figure 3: Main uses of smartphones in the representative sample of respondents. 1: email, web; 2: apps; 3: SMS, calls, no Internet; 4: music, movies, photos; 5: other things (Own work)

Three groups summing up to more than 60 % of valid answers are accessing Internet, using basic functions (SMS, calls), and a combination of Internet, applications, basic functions, and multimedia. The trend of smartphones complementing portable and stationary computers can be expected and clearly shows BYOD management in organizations is necessary to be addressed as security implications of compromised mobile devices will grow with more consumers replacing old technology with Internet-enabled endpoints.

However, it seems trust towards smartphones has not reached a point when they would be considered pertinent for sensitive online operations, e.g., banking: out of 746 valid answers, 469 participants (62.9 %) reported they would rather pick a PC/notebook instead of their mobile device; on the other hand, 243 (32.6 %) do not have issues with using smartphones in such circumstances. If the trend is indeed on the rise, we expect the uptake would be much slower especially in new users due to distrust towards and unfamiliarity with the new technology. In time, the attitude may change and accessing sensitive accounts may become a matter of habit, lowering inhibitions and security awareness in favor of comfort. Hence, countermeasures such as profiles should be deployed immediately after the device has been integrated into the organizational network.
The last question polled participants whether they would be willing to install a profile on their mobile device. Figure 4 demonstrates the results.

N	Valid	757
	Missing	27

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	I would mind	325	41,5	42,9	42,9
	No, I wouldn't mind	168	21,4	22,2	65,1
	Yes, as long as I can switch it off	133	17,0	17,6	82,7
	Only if the company wanted me to	131	16,7	17,3	100,0
	Total	757	96,6	100,0	
Missing	999	27	3,4		
Total		784	100,0		

Figure 4: Attitude of respondents towards profiles (Own work)

Around 43 % of respondents claimed profiles are unacceptable and dismissed them outright while 17.6 % were not against having a programmatic safeguards in place as long as they can be turned off should the need arise. Some users (22.2 %) do not see any issues with adding a profile regardless of whether it can be disabled. It is argued profiles are perceived as novel, unproven technology which coupled with lack of information about their underlying technical principles leads to misinformation. This can be counteracted by training and objective evaluation of advantages and disadvantages they pose for mobile devices, lessening enmity for profiles and BYOD management generally. Albeit opposition from some users will remain, the current situation would very likely change and the unacceptability rate decrease in favor of unconditional acceptance or assurance of administrative smartphone control retention. On the other hand, profiles are tailored for protecting sensitive organizational electronic assets with minimum impact to convenience; users should expect an adjustment period as new policies are implemented in day-to-day operations.

3.2 Passwords

Figure 5 presents length breakdown of the most important password the participants were asked to think of; the question did not specifically ask for the string.

N	Valid	771
	Missing	13

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<6 characters	143	18,2	18,5	18,5
	7--11 characters	435	55,5	56,4	75,0
	12--16 characters	155	19,8	20,1	95,1
	17--21 characters	23	2,9	3,0	98,1
	22+ characters	15	1,9	1,9	100,0
	Total	771	98,3	100,0	
Missing	999	13	1,7		
Total		784	100,0		

Figure 5: Password length breakdown (Own work)

Three out of four respondents in the sample have passwords no longer than 11 characters, with 56.4 % having their password between 7 and 11 characters. While the trend of moving away from shorter strings is clear, passwords have several attributes contributing to how resilient they are against reverse engineering attempts; these were discussed in Section 2.2. While emphasis is frequently put on length due to seemingly sound logic ("The longer the password, the harder it is to guess."), if the sequence is generic and predictable, length does not provide any security added value. Other aspects are therefore important as well. Lowercase and uppercase characters, numbers, and special symbols do not occur in real-world situations frequently due to decrease in user comfort when typing them. The following two figures demonstrate composition properties, specifically lowercase and uppercase characters (Figure 6), symbols and spaces (Figure 7). The participants were not asked for the number of letters to ensure their password could not be reconstructed in its entirety.

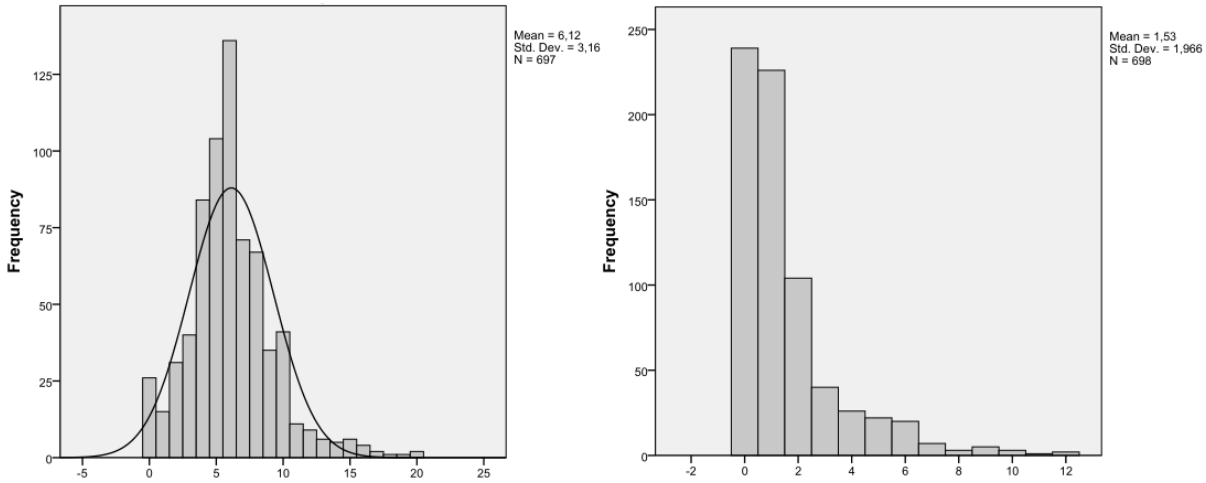

Figure 6: Distribution of lowercase (left) and uppercase (right) characters in passwords protecting respondents' most important account (Own work)

The results confirm users value comfort rather than security when composing and changing passwords. Out of 721 valid answers, the following was discovered:

- 83 (11.5 %) changed their password in last 30 days,
- 108 (15 %) did so during last three months,
- 121 (16.8 %) no more than 6 months back,
- 218 (30.2 %) during the last year,
- 186 (25.8 %) never changed their password,
- 5 (0.7 %) provided different answers usually stating they cannot remember exactly, suggesting more than a year.

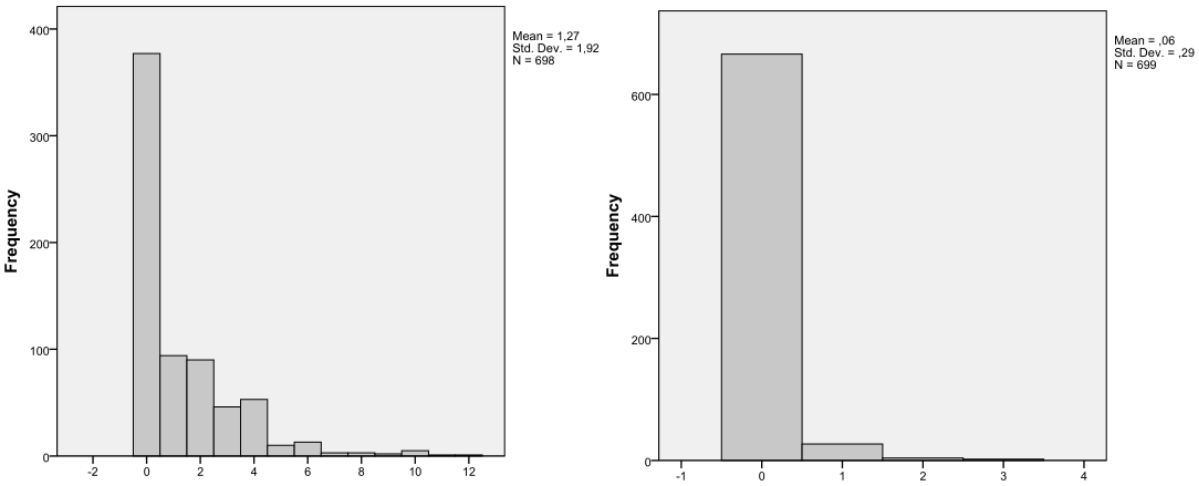

Figure 7: Distribution of special symbols (left) and spaces (right) in passwords protecting respondents' most important account (Own work)

Combination of weak passwords and their slow replacement creates viable attack vectors for adversaries. Should they get hold of hash-obfuscated strings, 56 % of users in the sample would use the same password for a year or more, and 72.8 % for at least 6 months, enough to use sophisticated reverse engineering techniques. Passwords predominantly include lowercase characters, 1—2 uppercase letters and symbols on average, and almost no spaces, limiting the search space considerably; uppercase letters can be with high probability expected on the very first position. These constraints allow to prune the candidate list further, decreasing the time factor involved. Moreover, more assumptions may hold which stems from total length. For example: while it is almost certain shorter (at most 6-character) passwords were selected for memorability, the same could likely be said about the 7—11 group, a length well within the capacity of an average user to remember, particularly when the string has personal significance. We hypothesize even very long strings could be defeated if suitable dictionary is employed.

Major composition rules identified include words found in dictionaries without modifications (248 out of 749 answers) and prepending/appending them with numbers, letters, or symbols (249); minor rules included typing random characters (117) and substitution (82). Both are trivial to implement in software and thus vulnerable. Plain dictionary entries only necessitates matching the correct word with the password, a fast algorithm executable in parallel even on consumer-grade setups. Depending on the quality of the prepended/appended sequence, the same could be stated of the second rule. Joining multiple common words together increases the time factor but should still be treated as unsafe because dedicated hardware components can substantially speed up the computations while not being overly taxing for a dedicated attacker. Substitution does little to thwart reverse engineering and due to its popularity among users, the rules have evolved and are now highly effective against common alterations. Typing random characters may seem secure; nevertheless, since the string needs to be memorized and no password manager is usually employed (see below), the characters likely represent keyboard patterns, e.g., "qwerty" guaranteed to be included in word lists freely available on the Internet.

The most popular choices for storing and reusing passwords are one master string for multiple accounts, and memorizing several unique credentials without reuse. The other two sparsely-populated categories are writing the password down and keeping it either private, e.g., wallet, or keeping it public, e.g., office desk. The secure alternative, password manager, was represented by 0.8 % of the 774 valid answers, a negligible proportion. Repercussions of a single string compromise includes complete electronic identity takeover, memorizing unique credentials also does not improve security as it is highly improbable an individual can recall multiple complex sequences of symbols sampled from all character sets organized so that they are impervious to dictionary and brute-force attacks. A more plausible scenario is several memorable passwords consisting of dictionary words combined together, or with a substitution scheme applied. While the threat is reduced to at most one account in this case, credentials management is cumbersome and concessions are likely made, e.g., incrementing the prepended/appended number and leaving the rest unchanged when rotating passwords. Writing down sensitive information is strongly discouraged as physical acquisition then becomes viable.

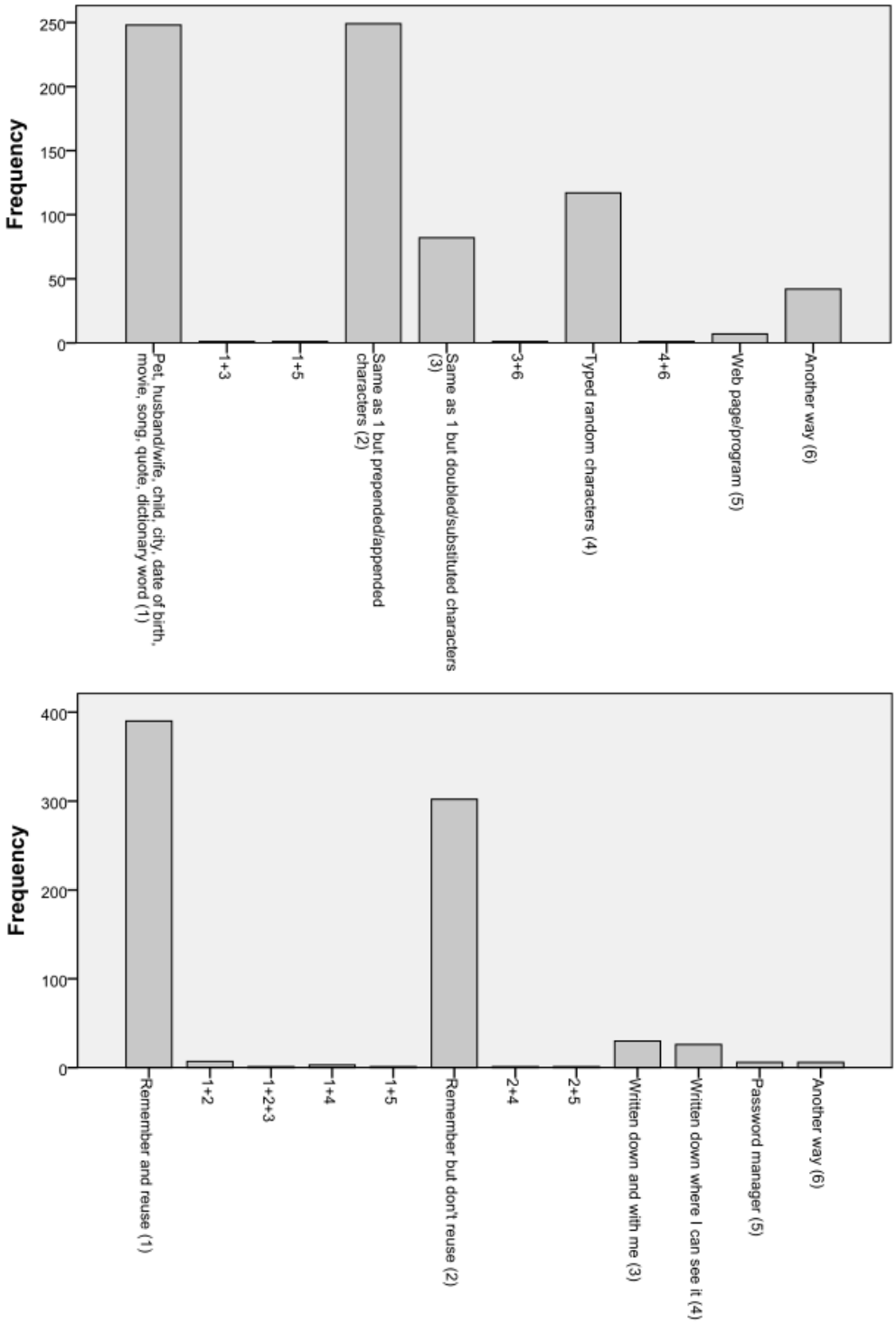

Figure 8: Selection rules for passwords (up) and their storage and reuse (down) in the representative sample (Own work)

4. Conclusion

The results are evidence which support conclusions made by other empirical studies in Section 2.2 claiming password composition policies are skewed towards memorability, with security of secondary importance. We believe the trend will continue until credentials managers do not gain traction among users. Organizations could pioneer the process and introduce enterprise-wide SECaaS solutions, i.e., provide security services which abstract from low-level technical details. By integrating ergonomics, graphical user interface (GUI) design, and human—computer interaction, users can be presented with attractive, clean, intuitive, and usable frontends.

Smart mobile devices are proliferated and some respondents utilize them for work-related tasks as well as entertainment. While desktop computers and notebooks boast larger screens and higher computational performance, broad application ecosystem, portability, and advanced features (Global Positioning System, touch interfaces) make smartphones a preferred choice for activities previously reserved for computers. With introduction of multi-core processing, we argue hardware performance will gradually become comparable to high-range stations while retaining energy efficiency. Tablets and phablets (portmanteau of "phone" and "tablets") address the limited screen estate problem and at the same time provide higher screen resolutions than average consumer-grade monitors. We are of the opinion users will increasingly tend towards mobile devices and security concerns in companies will have to be addressed accordingly.

Acknowledgments

The article was supported by the Centre for Security, Information and Advanced Technologies (CEBIA-Tech) project, registration number CZ.1.05/2.1.00/03.0089.

References

Ballagas, Raffael and Borchers, Jan and Rohs, Michael and Sheridan, Jennifer G. (2006) "The Smart Phone: A Ubiquitous Input Device", *IEEE Pervasive Computing*, Vol 5, No. 1, pp 70—77.

Friedewald, Michael and Raabe, Oliver. (2011) "Ubiquitous Computing: An Overview of Technology Impacts", *Telematics and Informatics*, Vol 28, No. 2, pp 55—65.

Gessner, Dennis and Giaro, Joao and Karame, Ghassan and Li, Wenting (2013) "Towards a User-Friendly Security-Enhancing BYOD Solution*", NEC Technical Journal*, Vol 7, No. 3, pp 113—116.

Keith W. Miller and Voas, Jeffrey and Hurlburt, George F. (2012) "BYOD: Security and Privacy Considerations", *IT Professional*, Vol 14, No. 5, pp 53—55.

Kelley, Patrick G. and Komanduri, Saranga and Mazurek, Michelle E. and Shay, Richard et al. (2012) "Guess Again and Again and Again: Measuring Password Strength by Simulating Password-Cracking Algorithms", *2012 IEEE Symposium on Security and Privacy*, May 20—23, 2012, San Francisco, California.

Kim, Ik-Soo. (2012) "Keypad Against Brute Force Attack on Smartphones", *IET Information Security*, Vol 6, No. 2, pp 71—76.

Klein, Daniel V. (1990) "'Foiling the Cracker': A Survey of, and Improvements to, Password Security", *2nd USENIX Security Workshop*, pp 5—14.

Malone, David and Maher, Kevin. (2012) "Investigating the Distribution of Password Choices", *21st International Conference on World Wide Web*, April 16—20, 2012, Lyon, France.

Sarga, Libor, and Jašek, Roman (2013) "Mobile Cyberwarfare Threats and Mitigations: An Overview", *12th European Conference on Cyber Warfare and Security (ECCWS-2013)*, Jyväskylä, Finland, July 11—12, 2013.

Scarfò, Antonio. (2012) "New Security Perspectives Around BYOD", *2012 Seventh International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA)*, Victoria, Canada, November 12—14, 2012.

Shirey, Robert W. (2007) "Internet Security Glossary, Version 2", [online], *Internet Engineering Task Force*, https://tools.ietf.org/html/rfc4949

Weiser, Mark. (1991) "The Computer for the 21st Century", *Scientific American*, Vol 265, No. 3, pp 94—104.

Yampolskiy, Roman V. (2006) "Analyzing User Password Selection Behavior for Reduction of Password Space", *2006 40th Annual IEEE International Carnahan Conferences Security Technology,* October 16—19, 2006, Lexington, Kentucky.

Yan, Jeff and Blackwell, Alan and Anderson, Ross and Grant, Alasdair. (2004) "Password Memorability and Security: Empirical Results", *IEEE Security & Privacy*, Vol 2, No. 5, pp 25—31.

Zviran, Moshe and Haga, William J. (1999) "Password Security: An Empirical Study", *Journal of Management and Information Systems*, Vol 15, No. 4, pp 161—185.

ECCWS 2015

Hatfield, UK

The 14th European Conference on Cyber Warfare & Security
University of Hertfordshire
Hatfield, UK
2-3 July 2015

9 781910 309247 >

For further information contact
info@academic-conferences.org
or telephone
+44-(0)-118-972-4148